L'interrogation philosophique
Collection dirigée par
Michel Meyer

Davidson et la philosophie du langage

PASCAL ENGEL

PRESSES UNIVERSITAIRES
DE FRANCE

A ma mère

ISBN 2 13 046416 5
ISSN 1159-6120
Dépôt légal - 1re édition : 1994, mai
© Presses Universitaires de France, 1994
108, boulevard Saint-Germain, 75006 Paris

« If you had liv'd, Sir,
Time enough to have been Interpreter
To Babell bricklayers, sure the Tower had stood. »

John Donne, *Satyre III*.

Sommaire

Introduction

Au début de l'exposition de son système des *Grundgezetze der Arithmetik* (1893), Frege soutient « que les conditions dans lesquelles [toute phrase de sa *Begriffsschrift*] désigne le Vrai sont déterminées à travers [ses] stipulations » et que « le sens de ce nom [d'une valeur de vérité] est la pensée, est le sens ou la pensée que ces conditions sont remplies » (§ 1.32). Dans le *Tractatus*, Wittgenstein écrit que « la proposition montre ce qu'il en est des états de choses *quand* elle est vraie. Et elle *dit* qu'il en est ainsi » (4.022). Au § 4.0424, il ajoute que « comprendre une proposition c'est savoir ce qui a lieu quand elle est vraie ». On dit souvent que la thèse selon laquelle la signification d'une phrase dans une langue naturelle consiste dans ses conditions de vérité est une banalité de la philosophie analytique du langage contemporaine. Mais on peut penser au contraire que depuis Frege et le *Tractatus*, elle s'est perdue. D'une part les positivistes l'ont écartée au bénéfice de l'idée que la signification doit être définie en termes des conditions de *vérification* des énoncés, et on s'est en général accordé sur le fait que l'identification de la signification et des conditions de vérité ne valait, à strictement parler, que pour les langues formelles, pour lesquelles Tarski et Carnap ont montré que l'on pouvait construire des définitions de la vérité. D'autre part les philosophes du « langage ordinaire » l'ont vigoureusement rejetée pour les langues naturelles, et ont vu dans les tentatives des logiciens pour appliquer à ces dernières les concepts

et les méthodes de la sémantique formelle au mieux des idéalisations illégitimes, et au pire des entreprises radicalement confuses. Ce que l'on peut appeler la conception *vériconditionnelle* de la signification — la signification est déterminée par les conditions de vérité — était donc tombée dans un relatif discrédit quand, dans les années 1960, le philosophe américain Donald Davidson l'a pour ainsi dire réinventée, sur des bases assez distinctes de celles qu'elle pouvait avoir chez Frege et ses successeurs. Davidson soutient en effet, à la différence de Tarski, que les méthodes par lesquelles celui-ci définit la vérité pour les langues formelles peuvent être, dans certaines conditions, appliquées aux langues naturelles, et servir à élucider la nature de la signification dans les langues naturelles. Ce qui motive principalement le projet de Davidson est d'abord l'idée, qu'il emprunte dans une large mesure à Quine, selon laquelle la notion même de signification, qu'elle soit ou non définie à partir de la notion fregéenne de *sens* (*Sinn*) ou à partir de la notion carnapienne d'*intension*, est obscure et confuse, et qu'il vaut mieux pour l'analyser partir des conditions dans lesquelles elle est attribuée à des expressions employées par des locuteurs dans diverses circonstances empiriques où ils donnent leur assentiment ou leur dissentiment à des phrases. C'est pourquoi Davidson propose d'aborder le phénomène de la signification dans les langues naturelles à partir de ce qu'il appelle des *théories* de la signification, et d'envisager la *forme* que doivent prendre ces théories. Puisque le concept même de signification ne peut selon lui être supposé donné ou compris d'avance, on doit aborder la construction d'une théorie de la signification à partir d'un autre concept, à la fois plus primitif, plus simple et plus adéquat empiriquement, celui de vérité. C'est ici que s'introduit d'abord la conception vériconditionnelle. On supposera que les locuteurs connaissent les conditions de vérité des phrases de leur langage, et que cette connaissance est au moins en partie une composante de leur connaissance de ce que ces phrases signifient. On supposera aussi qu'ils ont diverses attitudes, repérables empiriquement, vis-à-vis des phrases qu'ils tiennent pour vraies. Et on essaiera d'interpréter leur langage sur la base de ces données minimales. La théorie vériconditionnelle de la signification intervient alors à un autre titre. Comme l'avait souligné Wittgenstein (*Tractatus*, 4.027), « Il est dans la nature de la proposition de pouvoir nous communiquer

un sens *nouveau* », notamment parce qu'« on la comprend quand on comprend ses constituants » (4.024). Les phrases sont composées de parties, et leur sens est déterminé par le sens de ces parties. Elles ont, en d'autres termes, une structure *compositionnelle*, comme l'avait vu Frege. Cette structure compositionnelle explique que l'on puisse former à partir d'expressions données des phrases nouvelles, et les comprendre. Or on peut établir cette structure comme affectant précisément les conditions de vérité des phrases. C'est parce qu'elles ont la structure sémantique qu'elles ont que les phrases peuvent avoir des conditions de vérité déterminées. Une théorie de la vérité déterminera cette structure. Davidson propose d'employer à cette fin le type de structure postulée par les théories de la vérité employées par Tarski pour les langues formelles. Pour discerner une structure dans le langage du locuteur, l'interprète construit pour ce langage une théorie de type tarskien, destinée à montrer comment les phrases sont composées à partir d'un ensemble d'éléments finis. On pourra ainsi expliquer comment un ensemble infini de phrases peuvent être apprises et comprises, et de proche en proche déterminer la signification de l'ensemble des phrases du langage. L'un des principes de base de la méthode de Davidson est que ces attributions de signification ne peuvent pas s'effectuer phrase par phrase, expression par expression, mais sur un ensemble de phrases tenues pour vraies par les locuteurs, selon une méthode holistique. L'interprétation du langage est holistique en un autre sens, car elle est étroitement liée à celle des croyances et des autres états psychologiques que les phrases expriment. Une théorie de la signification devra s'intégrer au sein d'une théorie générale de l'interprétation du comportement et des actions humaines. Cette théorie ne s'appuie pas seulement sur une théorie sémantique, mais aussi sur des principes normatifs de rationalité, et en particulier le « principe de charité », qui prescrit de tenir les croyances et les attitudes psychologiques des agents comme vraies et cohérentes pour la plupart. C'est parce qu'il met l'accent sur la notion de vérité, et non pas sur celle de signification, ou plutôt parce qu'il entend élucider la première à partir de la seconde, que Davidson doit supposer que les conditions de vérité des phrases d'un langage sont déjà largement comprises par ceux qui le parlent.

L'approche de Davidson a d'importantes affinités avec la manière dont un certain nombre de logiciens et de linguistes ont conçu la tâche d'une

sémantique des langues naturelles dans les années 1960. Comme celle de Chomsky, elle vise à rendre compte du caractère productif du langage humain, et elle s'apparente de prime abord à celle des logiciens qui, à la suite de Montague, ont cherché à appliquer des méthodes formelles à l'étude des langues naturelles. Mais à la différence de Chomsky, Davidson n'entend pas proposer d'hypothèses psychologiques sur la compétence linguistique. Il voit plutôt la signification comme le produit de la communication entre des locuteurs, et la compréhension du langage comme une forme d'interprétation, procédant, d'un point de vue extérieur aux propriétés psychologiques des locuteurs, à des attributions « à la troisième personne ». A la différence de Montague, Davidson n'entreprend pas de construire dans le détail une théorie sémantique logique applicable aux langues naturelles, et il refuse le cadre « intensionnaliste » de la plupart des travaux récents dans ce domaine, parce qu'il pense qu'une théorie empirique de l'interprétation du langage doit se fonder sur des concepts extensionnels, présupposant la machinerie logique la plus minimale possible. Comme les théoriciens de la « sémantique générative » Davidson considère qu'une sémantique pour les langues naturelles doit s'appuyer sur une théorie de la « forme logique » des phrases, de manière à révéler leurs structures, et il a proposé selon ces principes plusieurs analyses très discutées de la forme logique des phrases d'action, des phrases adverbiales et des phrases rapportant des contenus d'attitudes propositionnelles. Mais l'objectif de Davidson n'est pas principalement linguistique ou sémantique. Il est philosophique et critique, et s'inscrit, comme l'a souligné Crispin Wright (1987), dans la lignée des grands projets de « reconstruction rationnelle » auxquels la philosophie analytique du vingtième siècle nous a rendus familiers, comme le projet de fonder rationnellement nos inférences inductives, ou le projet de fonder le langage de la science sur une base physicaliste, ou — pour prendre des exemples qui indiquent que ce genre de tentative est loin d'avoir disparu dans la philosophie contemporaine — le projet d'une « naturalisation » de l'intentionnalité et des phénomènes mentaux, le projet de reconstruire les mathématiques et la physique sur des bases strictement nominalistes, ou encore celui de fournir une interprétation empiriste et instrumentaliste des sciences. La comparaison entre le projet davidsonien et ces divers programmes sera sans doute propre

à susciter le scepticisme et la méfiance de tous ceux qui sont convaincus qu'ils sont voués à l'échec, et qu'on peut montrer qu'ils le sont. Mais il y a cependant une différence importante : contrairement à nombre de projets « naturalistes » contemporains, celui de Davidson ne part pas d'une hypothèse réductionniste. Davidson ne soutient pas que la notion de signification pourrait être analysée en des termes plus primitifs qui ne fassent pas appel à des notions sémantiques ou intentionnelles. Il n'entreprend pas de fonder la sémantique sur la psychologie, sur la biologie, ou sur des bases physicalistes. Il n'en conclut pas pour autant qu'aucune théorie systématique de la signification n'est possible, et en ce sens il s'oppose aussi bien à diverses formes de scepticisme quant à la possibilité d'une telle théorie qu'à diverses formes de conceptions « herméneutiques » qui interdiraient toute forme de confirmation empirique de nos attributions de signification ou d'états mentaux. Il partage cependant avec ces positions l'idée que notre mode d'accès aux significations et aux contenus intentionnels est fondamentalement distinct de celui que nous avons aux phénomènes décrits par les sciences de la nature.

Bien qu'on ait beaucoup discuté, depuis la parution en 1967 de « Truth and Meaning », ce que l'on a appelé « le programme de Davidson » dans la perspective d'une sémantique, et que les diverses analyses de la « forme logique » de fragments de langue naturelle, telles que les phrases d'action, les modifications adverbiales, les termes de masse, les quantificateurs ou les descriptions définies aient occupé les efforts de nombreux philosophes, linguistes et logiciens, ce n'est pas sur cet aspect des choses que porte principalement ce livre, mais sur le projet même d'une théorie de la signification et sur la philosophie du langage qui l'inspire. La question que j'ai posée ici est celle que Wright (1987) pose ainsi : « Comment la théorie de la signification peut-elle être un projet philosophique ? » Cela implique qu'on soit en mesure non seulement de déterminer ce que des analyses comme celles de Davidson ont à voir, ou n'ont pas à voir, avec des analyses linguistiques, psychologiques, ou neuropsychologiques de la compétence que les locuteurs manifestent quand ils parlent un langage, mais aussi que l'on cerne quelles sont les implications proprement philosophiques de ce projet. De ce point de vue, la question de savoir quelle forme pourrait prendre une théorie de la signification pour une langue

naturelle ne prend pas tant place dans le cadre d'une explication, scientifique ou non, des phénomènes liés à ce que l'on appelle « signification », que dans le cadre métaphysique d'une analyse des liens entre le langage et la réalité, et des liens entre notre connaissance du langage et notre connaissance de la réalité. Selon Michael Dummett, Frege a définitivement établi un certain style de questions et de méthodes en philosophie, en soumettant l'analyse du contenu des pensées, en tant que contenus objectifs des jugements (et non pas de représentations mentales) à une analyse du langage et de l'expression linguistique des pensées. La question de savoir comment une représentation objective de la réalité est possible est devenue, chez les philosophes analytiques, la question de savoir comment des phrases peuvent avoir une signification, et en quoi elles peuvent désigner une réalité objective. En d'autres termes, la théorie de la connaissance et la métaphysique se sont trouvées soumises à la philosophie du langage, promue par là au rang de « philosophie première » et de méthode philosophique privilégiée. Expliciter les principes systématiques de notre langage, c'est par là même déterminer comment nous pouvons saisir une réalité objective. Selon Dummett, une théorie de la signification a pour but d'établir ces principes, et en ce sens le projet davidsonien s'inscrit explicitement dans la tradition fregéenne. Mais précisément parce qu'une théorie de la signification doit être, pour Dummett, une théorie de ce que nous comprenons, quand nous comprenons un langage, elle doit satisfaire une condition que l'analyse fregéenne de la signification en termes de conditions de vérité ne satisfait pas, si ces conditions de vérité sont supposées indépendantes de la manière dont nous pouvons les connaître. Comment, en effet, pourrions-nous connaître et manifester la signification des phrases de notre langage si cette signification est déterminée par des conditions de vérité qui peuvent être, par principe, inconnues de nous ? Dummett appelle *réaliste* une telle conception de la signification, l'assimile dans une large mesure à la position de Davidson et en critique les hypothèses fondamentales : le principe vériconditionnel, le holisme selon lequel le sens d'une phrase ou d'une expression dépend du sens d'autres phrases ou expressions, et l'hypothèse d'une précompréhension de la notion de vérité qui rendrait possible la démarche minimaliste de Davidson, qui cherche à éviter tout recours à la notion de signification dans les principes

de la théorie. Il lui oppose une conception *antiréaliste*, d'après laquelle la signification doit s'analyser en termes de conditions d'assertion des phrases, et qui rejette les postulats holistes et minimalistes au bénéfice d'une explication constructive du sens. Le débat entre le « réalisme » et « l'antiréalisme » en sémantique a, au cours des vingt dernières années, pris une ampleur considérable, en grande partie parce qu'il semble promettre, sur des bases proprement sémantiques, une voie d'approche de certaines des questions métaphysiques et épistémologiques les plus fondamentales. Mais les choses sont vite apparues beaucoup plus complexes. Dummett soutient que le réalisme des conditions de vérité ne peut fonder une théorie de la signification parce qu'il ignore certaines contraintes épistémiques qui pèsent sur toute théorie de ce type : que le sens soit manifestable dans l'usage, qu'il n'y ait pas plus dans le sens que nous ne sommes capables de comprendre. Mais la nature exacte de ces contraintes, telles que les formule l'antiréalisme, n'est pas évidente, et il n'est pas certain non plus qu'on doive en conclure, comme Dummett, que la notion de vérité elle-même doive devenir elle aussi une notion épistémique, et qu'il faille réviser la logique classique. Si c'est le cas, les implications métaphysiques que peut avoir le débat réalisme/antiréalisme en sémantique seront loin d'être aussi claires qu'elles peuvent apparaître de prime abord.

J'ai voulu, dans ce livre, examiner ces débats, et tenter de proposer un cadre d'analyse qui permette de les arbitrer. J'ai d'abord cherché à présenter les thèses principales de Davidson en philosophie du langage. C'est à cette présentation que sont consacrés les trois premiers chapitres. Dans le premier, j'expose sa conception de la forme d'une théorie de la signification, et sa démarche « indirecte » qui entreprend d'extraire la signification à partir des conditions de vérité. Contrairement à une idée répandue, Davidson n'identifie pas ces deux notions : il soutient plutôt que, si l'on respecte certaines conditions formelles et empiriques, une théorie de la vérité pourra faire office de théorie de la signification. En ce sens, le projet d'une théorie de la signification fait partie d'un cadre plus large, celui d'une théorie de l'interprétation « radicale », destinée à assigner à la fois des contenus sémantiques à des phrases et des contenus intentionnels à des croyances, à partir d'une interprétation des actions. J'expose les principes de cette conception de l'interprétation dans le chapitre 2.

Parce qu'elle est étroitement liée à la philosophie de l'action et du mental de Davidson (1980), il est nécessaire de ne pas la dissocier des thèses de Davidson dans ces domaines. Mais je n'ai pas prétendu ici examiner ces thèses pour elles-mêmes, bien que je sois à de nombreuses reprises amené à indiquer les points de contact. C'est pourquoi on ne trouvera pas, ou peu, dans ce livre, de traitement systématique de certains des problèmes les plus discutés dans la philosophie de l'esprit contemporaine, comme celui de l'externalisme ou de l'individualisme au sujet des contenus intentionnels, bien que ces problèmes affleurent sans cesse, et que je ne parvienne pas à me convaincre qu'on puisse, en philosophie, faire une seule chose à la fois. Dans le chapitre 3, j'aborde la question de savoir comment peuvent s'articuler, chez Davidson, sémantique et pragmatique, et notamment comment une théorie essentiellement vériconditionnelle de la signification peut s'articuler avec une théorie de la signification en termes d'intentions des locuteurs, comme celle de Grice. La réponse est à nouveau à trouver dans les conditions de l'interprétation. Les quatre autres chapitres du livre sont axés autour du débat réalisme/antiréalisme. Le quatrième expose les arguments de Dummett contre les théories réalistes de la signification, et en particulier celle de Davidson. Dans le chapitre 5, j'évalue la portée de ces critiques, admets qu'une conception réaliste doit poser certaines conditions épistémiques quant à la signification, mais rejette l'inférence que l'antiréaliste en tire quant à la nature de la vérité : celle-ci ne se laisse pas réduire à l'assertabilité. A partir d'une analyse de la notion de vérité, je soutiens que la conception vériconditionnelle est fondée sur des platitudes que l'antiréaliste n'a pas besoin de rejeter, et qu'on peut formuler une position de type réaliste, qui à la fois se distingue du réalisme métaphysique ou « externe », répond à certaines conditions antiréalistes, et satisfait les critères d'objectivité du sens et de la vérité sur lesquels repose toute conception réaliste. J'appelle cette conception « réalisme minimal ». Dans le chapitre 6, je soutiens que Davidson est un réaliste minimal, mais que son holisme menace cette thèse réaliste. Je m'efforce alors de formuler une version acceptable du holisme, qui conduit à abandonner certains des principes de base de la théorie davidsonienne de l'interprétation. Dans le septième et dernier chapitre, je m'adresse à la question fondamentale posée à la fois par Davidson et par Dummett : en quel sens

une théorie systématique de la signification peut-elle être une théorie de la compréhension du langage ? Je soutiens que la position de Davidson ne permet de répondre que partiellement à cette question, parce qu'elle exclut toute analyse détaillée de la compétence effective des locuteurs. Pour cela on doit formuler une conception de la « connaissance tacite » du langage s'appuyant sur des bases psychologiques. Mais cette entreprise se heurte à certaines questions qu'on peut tirer des considérations de Wittgenstein sur la notion de règle. La critique wittgensteinienne peut prendre la forme d'un scepticisme radical quant à la possibilité même d'une théorie systématique de la signification, tel que l'a proposé Kripke (1981). Tout en prenant acte de ces critiques, je m'efforce de montrer qu'une théorie de la compétence sémantique s'appuyant sur la notion de connaissance tacite est possible.

Bien que ce livre soit centré sur un exposé de la philosophie du langage de Davidson, j'ai également voulu dire sur quels points j'étais en désaccord avec lui. Mais tout désaccord suppose une compréhension et un accord. Davidson n'est pas un auteur facile : il distille ses idées dans des articles élégants et laconiques, qui ont souvent inspiré des malentendus. C'est pourquoi j'ai été amené, peut-être plus que je ne l'aurais souhaité, à consacrer beaucoup de place à exposer ses thèses, avant d'envisager les tensions et les difficultés auxquelles elles me paraissaient conduire. On trouvera peut-être que j'abuse souvent, dans ce livre et ailleurs, des solutions « modestes » ou « minimalistes » à des débats qui semblent appeler des confrontations radicales et tranchées. Mais il n'y a là de ma part aucun souci de conciliation ou d'aplatissement des positions en présence. C'est précisément parce que les phénomènes comme ceux dont il est question ici — la nature de la signification, des contenus mentaux et de l'intentionnalité — sont à la fois systématisables et fuyants, dépendants de conditions naturelles et de conditions normatives, qu'une position « minimaliste » comme celle défendue ici est difficile et, je l'espère, contestable.

Remerciements

Cet ouvrage a pour ancêtre une thèse de doctorat d'Etat soutenue à l'Université de Provence en octobre 1990, et s'appuie sur des travaux plus anciens. Il se veut, en un sens, une enquête complémentaire, en philosophie du langage, à celle que j'avais menée en philosophie de la logique dans *La norme du vrai*. J'espère que seuls les développements les plus aptes ont survécu à la révision, même si on a pu, et pourrait sûrement encore, souhaiter que la pression sélective ait été plus forte. J'ai en particulier supprimé une longue analyse de la notion de forme logique à travers l'analyse davidsonienne des phrases d'action et des modifications adverbiales. Parmi les membres de mon jury, je remercie d'abord Gilles Granger, mon directeur, pour son soutien et son encouragement en plus d'une occasion. J'ai adopté des orientations et des méthodes en philosophie du langage qui divergent des siennes, mais je n'ai jamais cessé d'être inspiré par les questions qu'il a posées dans son œuvre. Je remercie également Jean-Claude Pariente pour ses critiques attentives, et tout particulièrement Thomas Baldwin, dont l'amitié, la connaissance profonde des sujets traités ici et les objections lucides m'ont été d'un grand secours (je ne suis pas sûr d'avoir répondu à ces objections). La dette que j'ai envers Jacques Bouveresse est profonde et ancienne, et je lui suis très reconnaissant pour son appui amical dans ces circonstances, alors que j'étais tenté de céder à un certain découragement (qui n'avait rien à voir avec le scepticisme analysé au chapitre 7 de ce livre). Je dois aussi beaucoup à Daniel Laurier et à Michel Seymour, dont les écrits sur ces sujets sont pour moi un modèle. Pour m'avoir encouragé et aidé en diverses occasions, je remercie aussi Per Aage Brandt, François Clementz, Michael Dummett, Paul Gochet, Pierre Livet, Frédéric Nef, Jean-Luc Marion, Christopher Peacocke, et Claudine Tiercelin. Enfin, je voudrais exprimer ma profonde reconnaissance à Donald Davidson. Depuis qu'il m'a invité, en 1985, à lui rendre visite à Oxford, il ne m'a pas ménagé son

amitié et ses conseils. J'espère néanmoins que ce livre n'est pas un témoignage autobiographique d'admiration pour son oeuvre.

Les références aux essais de Davidson sont données par leurs dates de publication d'après la liste qui figure en bibliographie (dans le style : « Davidson 1966 a »). Quand ils sont repris dans *Inquiries into Truth and Interpretation* (1984) et dans *Essays on Actions and Events* (1985), les numéros de pages sont donnés d'après ces éditions, et sont suivis, en italiques, de la pagination de mes traductions françaises de ces ouvrages, *Enquêtes sur la vérité et l'interprétation*, Nîmes, J. Chambon, 1993, et *Actions et événements*, Paris, PUF, 1993 (dans le style « Davidson, 1967 : 23, *49* » ou quelquefois seulement « 1967 : 23, *49* »). Je n'ai pu résister à la tentation d'emprunter à Blackburn (1984) la citation de Wodhouse mise en exergue du chapitre 2. Quant à celle du chapitre 7, je l'ai empruntée à Dummett.

Les technicités ont été réduites au minimum. Je me suis permis souvent, dans le schéma de Tarski : « S » est vrai ssi *p*, d'écrire simplement « *S* est vrai ssi *p* », où « ssi » abrège « si et seulement si ».

I

Théories de la vérité et théories de la signification

Le professeur. — Comment dites-vous, par exemple, en français : « Les roses de ma grand-mère sont aussi jaunes que mon grand-père qui était asiatique ? »...
L'élève. — Eh bien, on dira, je crois : les roses de ma... comment dit-on grand-mère, en français ?
Le professeur. — En français ? Grand-mère.
L'élève. — Les roses de ma grand-mère sont aussi... jaunes, en français, ça se dit « jaunes » ?
Le professeur. — Oui évidemment !
L'élève. — Sont aussi jaunes que mon grand-père quand il se mettait en colère.
Le Professeur. — Non... qui était a...
L'élève. — ...siatique...
Le professeur. — C'est cela. »

E. Ionesco, *La leçon.*

1.1. *Qu'est-ce qu'une « théorie de la signification » ?*

Dans « Truth and Meaning » (1967) et dans une série d'essais, Davidson pose la question de savoir quelle forme devrait prendre une théorie de la signification pour les langues naturelles, et propose une réponse à cette question qu'on a appelée « programme de Davidson ». Mais la nature même de la question et du programme proposé ne sont pas immédiatement clairs, et nombre de malentendus peuvent surgir quand il s'agit de savoir ce que Davidson entreprend réellement. On tentera donc d'abord de clarifier la nature du projet.

Tout d'abord, la question : « Quelle forme devrait prendre une théorie de la signification ? » est ambiguë, parce que l'expression « théorie de la signification » peut avoir au moins deux sens distincts.

1 / En un premier sens, une « théorie de la signification » est une spécification détaillée de la signification de toutes les phrases d'une langue naturelle. A chaque phrase *s* d'une langue L, une telle théorie assignera une signification, au moyen d'une description de la forme :

s signifie (dans *L*) que *p*

où la phrase « *p* » sera dite « donner la signification » de *s*. En d'autres termes, une théorie de la signification sera ce que l'on appelle couramment une sémantique pour *L*, et à chaque phrase de *L* elle attribuera une représentation sémantique de cette phrase.

2 / En un second sens, plus général, une « théorie de la signification » est une analyse du concept de signification, une théorie philosophique de la signification, donnant une définition ou une explication de ce concept, ou tout au moins le situant dans un ensemble d'autres concepts. Une théorie de la signification, en ce sens, fournirait des réponses à des questions comme celles-ci : y a-t-il des entités ou des choses telles que les significations ? quelle serait leur nature et leurs conditions d'identité ? le concept de signification peut-il être défini ou éclairé par des concepts voisins, tels que ceux de synonymie, de traduction, ou d'analyticité ?

Il est assez naturel de supposer que ces deux sens de l'expression « théorie de la signification » correspondent à une division du travail assez familière. La tâche de construire des théories de la signification au premier sens semble relever de la compétence du linguiste, et faire partie d'une entreprise de description de la structure des langues naturelles. La tâche de construction d'une théorie de la signification au second sens semble relever de la compétence du philosophe. Il y a bien des manières d'envisager la forme que pourrait prendre une théorie de la signification en l'un ou en l'autre sens. Dans le cadre des théories linguistiques, il existe toutes sortes de façons de décrire les faits que l'on peut tenir comme relevant de la signification, à commencer par une classification de ces faits au sein de l'une ou l'autre des disciplines qui composent une sémiotique, selon la tripartition de Morris : syntaxe, sémantique, et pragmatique. Il est également banal de constater qu'en philosophie, la notion de signification est susceptible d'analyses très diverses. Par exemple une théorie de la signification au second sens peut prendre la forme d'une « explication »,

au sens carnapien, de cette notion, en termes d'intension et d'extension (Carnap, 1956). On peut chercher à fournir une réduction béhavioriste de la signification en termes de dispositions au comportement, ou une autre forme de réduction physicaliste ou naturaliste. On peut encore, comme Quine, abandonner une tentative de définition de la notion de signification, et entreprendre plutôt d'analyser les conditions de la formulation de manuels de traduction d'un langage dans un autre.

Pour chacun des sens de l'expression « théorie de la signification », la question de savoir quelle forme pourrait prendre respectivement une théorie satisfaisante est donc pleinement justifiée. Il serait approprié d'adopter une convention reçue, et de parler, pour distinguer plus nettement ces deux sens, d'une « TS » pour désigner une théorie sémantique pour une langue naturelle, et d'une « théorie de la signification » pour désigner une analyse philosophique du concept de signification (Davies, 1981, Peacocke, 1985). Mais le fait que nous devions distinguer ces deux types de théories n'implique pas qu'elles soient indépendantes. Le but de Davidson n'est pas simplement de formuler des critères d'adéquation d'une sémantique linguistique d'un côté, et de formuler une théorie philosophique de la signification de l'autre. Son but est d'apporter une réponse à la seconde question en apportant une réponse à la première. En d'autres termes, répondre à la question :

(i) quelle forme devrait prendre une théorie qui spécifierait la signification de toutes les phrases d'une langue naturelle (une TS) ?

est une manière de fournir une théorie philosophique de la signification. Dummett a, sur ce point, bien caractérisé la position de Davidson :

> [Selon Davidson], la meilleure méthode pour formuler les problèmes philosophiques attenants à la notion de signification et aux notions voisines est de demander quelle forme devrait être prise par ce que l'on appelle une « théorie de la signification » pour un langage tout entier ; c'est-à-dire une spécification détaillée des significations de tous les mots et des opérateurs qui servent à former des phrases du langage, de manière à produire une spécification de toute expression et phrase du langage. Ce n'est pas que la construction d'une théorie de la signification, en ce sens, pour un langage, soit considérée comme un projet

> pratiquement réalisable ; mais on pense que, une fois que nous pouvons énoncer les principes généraux en accord avec lesquels une telle construction peut être accomplie, nous serons arrivés à une solution des problèmes concernant la signification qui suscitent la perplexité des philosophes (Dummett, 1975, 97 ; cf. aussi Foster, 1976, 4).

L'objectif de Davidson n'est donc pas principalement linguistique. Il est philosophique : il entend déterminer ce que serait une TS, afin d'éclairer en retour la notion de signification. Pourquoi cette démarche est-elle appropriée, s'agissant de cette notion ? C'est précisément parce que nous n'avons *a priori* aucune idée de ce que peut être une théorie de la signification que nous pouvons espérer élucider cette notion en nous interrogeant sur la forme d'une TS. Davidson propose donc une approche indirecte du concept de signification à travers l'examen des TS.

Cette stratégie indirecte ne risque-t-elle pas d'être circulaire ? Car comment pourrons-nous déterminer qu'une TS a une forme satisfaisante, si nous ne savons pas de quoi elle est la théorie, si nous n'avons au préalable aucune idée de ce qu'est, pour une phrase d'un langage, qu'une bonne attribution de signification ? N'avons-nous pas besoin de déterminer quels faits, s'il y en a, sont des faits de signification, et par conséquent d'avoir au préalable quelque chose comme une théorie ou une analyse du concept même de signification ? Nous avons en particulier besoin de répondre à une autre question que (i), qui serait la suivante :

(ii) en vertu de quoi une TS plutôt qu'une autre est-elle applicable aux faits que l'on peut tenir pour caractéristiques de la signification pour un langage compris et utilisé par une communauté donnée ?

Cette question n'est pas la même que la précédente, parce qu'on peut très bien s'accorder sur une réponse à la première question, sans s'accorder sur une réponse à la seconde. En d'autres termes, on peut admettre une conception commune de la forme générale d'une sémantique pour une langue, sans pour autant admettre que la fonction d'une sémantique est d'expliquer un seul et même type de faits. Bien qu'il tienne la notion de signification pour obscure, Davidson n'en conclut pas pour autant qu'il n'y ait pas de faits de signification et que ces faits ne soient pas déterminables.

Il admet que sa stratégie indirecte ne revient pas à faire l'économie d'une analyse préalable de la notion de signification, et qu'en ce sens une réponse à (i) ne peut pas se substituer complètement à une réponse à (ii). Mais cette stratégie s'efforce d'identifier les réquisits minimaux d'une analyse de la signification sans faire appel *directement* à cette notion ou à d'autres notions voisines. Il s'agira donc d'identifier un ensemble de conditions nous permettant de dire que telle ou telle TS joue adéquatement son rôle de spécification des significations des phrases d'une langue donnée sans présupposer une définition explicite de cette notion. Cette stratégie minimaliste n'implique en rien que la réponse à la question (i) ne fasse pas appel, implicitement, à une certaine théorie de la signification, ni que Davidson entende éliminer radicalement tout usage de cette notion en philosophie du langage. Au contraire, comme on le verra, il pense que le concept de signification est étroitement lié à d'autres, comme ceux de croyance et d'intention, et qu'il ne peut pas être réduit à des notions primitives ni éliminé.

Pour ces raisons, nous pouvons soupçonner que la différence entre une TS et une théorie de la signification ne sera pas aussi grande que ce que sa méthode ne le suggère, c'est-à-dire que les principes qui auront conduit à la construction d'une théorie sémantique satisfaisante reviendront bien à proposer une conception et une analyse générale de la signification[1]. On gardera néanmoins à l'esprit que l'assimilation plus ou moins tacite qui en résultera entre l'énoncé d'une méthode pour décrire

1. La distinction entre une TS et une théorie de la signification a surtout un intérêt méthodologique, comme le montre le fait que Davidson lui-même ne tient pas à séparer la théorie des conditions formelles d'une théorie de la signification de la théorie de ses conditions empiriques :

« Théorie de la signification » n'est pas un terme technique, mais un geste fait en direction d'une famille de problèmes (un problème famille). Parmi les problèmes se tient centralement la tâche d'expliquer le langage et la communication en faisant appel à des concepts plus simples, ou en tout cas différents. Il est naturel de croire que c'est possible parce que les phénomènes linguistiques sont si évidemment survenants sur des phénomènes non linguistiques. Je propose d'appeler une théorie « théorie de la signification » pour une langue naturelle L si elle est telle que (a) la connaissance de la théorie suffit pour comprendre les énoncés des locuteurs de L et (b) la théorie peut recevoir une application empirique en recourant à des données décrites sans utiliser de concepts inguistiques, ou tout au moins sans utiliser des concepts sémantiques qui soient spécifiques aux phrases et aux mots de L. La première condition indique la nature de la question ; la seconde requiert qu'elle ne soit pas fondée sur une pétition de principe » (1977 : 215, 312).

la signification des phrases d'un langage et une analyse philosophique de la signification ne sera justifiée en dernière instance que si les principes et les conditions générales stipulés par Davidson pour la construction d'une telle théorie sont eux-mêmes justifiés.

Quels sont ces principes et ces conditions ? Dans les écrits de Davidson, on trouve trois sortes de conditions d'adéquation pour la construction de théories de la signification, qu'on appellera respectivement des conditions *constitutives*, des conditions *formelles*, et des conditions *empiriques*. Les conditions constitutives sont celles qui justifient le projet même de construction d'une théorie de la signification. Elles découlent de deux faits apparemment incontestables : les locuteurs d'une langue naturelle comprennent leur langage, et ils sont en mesure, sur la base des énonciations d'autres locuteurs, d'interpréter ce que disent ces locuteurs. Comprendre une expression, c'est savoir ce qu'elle signifie : en ce sens, une théorie de la signification est une théorie de la compréhension ou de la connaissance du langage. L'hypothèse de base est donc que les locuteurs comprennent et interprètent leur langage en vertu d'une certaine connaissance, et qu'il est possible d'établir quelle sorte de connaissance est responsable de cette capacité de compréhension et d'interprétation. Comme on le verra, le but de Davidson n'est pas d'expliquer cette capacité au sens où pourrait le faire une théorie psychologique ou neuropsychologique de la compréhension du langage. La question n'est pas de savoir ce que nous connaissons en fait et qui nous permet de comprendre et d'interpréter un langage, mais de savoir ce qui pourrait suffire pour que nous comprenions et interprétions un langage. De ces conditions constitutives découlent des conditions formelles. Une TS pour une langue naturelle a une certaine structure, et cette structure est, dans une large mesure, comparable à celle des théories sémantiques que les logiciens construisent pour les langues formelles. Mais pour qu'une théorie de la signification réponde aux conditions constitutives, elle doit satisfaire certains critères formels bien précis. Ces critères concernent la nature des énoncés qui, dans une théorie de la signification sont supposés spécifier la signification des phrases, la distinction entre ceux qui donnent la signification d'expressions primitives du langage et ceux qui donnent celle d'expressions dérivées, ou encore la nature des concepts qui doivent figurer dans ces énoncés.

Enfin, Davidson impose à une TS certaines conditions empiriques. Une TS doit pouvoir être testable, c'est-à-dire se prêter à des attributions vérifiables de significations aux locuteurs d'une langue et d'une communauté données. Cela n'implique pas que Davidson considère qu'une théorie sémantique en général soit une théorie empirique, au même titre que des théories physiques ou chimiques, ni que la méthodologie d'ensemble qu'il adopte soit béhavioriste. Au contraire, l'une des thèses fondamentales de Davidson est que toute attribution aux énonciations d'un locuteur de certaines significations présuppose l'emploi de certains principes normatifs de rationalité qui ne sont pas testables empiriquement, et il rejette explicitement le béhaviorisme. Ce que les conditions empiriques d'une TS impliquent en revanche est que la validité d'une telle théorie ne se mesure pas simplement, comme c'est souvent le cas pour les théories sémantiques, aux intuitions particulières que les locuteurs ont de la signification des expressions de leur langage, mais qu'il y ait certaines bases objectives à partir desquelles on puisse effectuer des attributions de signification. L'ensemble de ces conditions empiriques fait l'objet de ce que Davidson appelle « théorie de l'interprétation radicale ».

Ces conditions doivent nous donner une image de ce que serait une TS pour une langue naturelle, et il n'est pas possible de les isoler les unes des autres. Aucune d'entre elles n'est suffisante. Par exemple des théoriciens pourraient s'accorder sur certaines des conditions formelles, sans que pour autant les conditions empiriques soient satisfaites. Nombre des objections adressées à Davidson viennent de ce que l'on a isolé un groupe de critères au détriment des autres. Ses écrits se divisent assez naturellement en deux groupes : ceux dans lesquels il analyse les réquisits formels des TS, et ceux dans lesquels il se concentre plus directement sur les réquisits empiriques. Les premiers portent essentiellement sur le rôle joué par une théorie de la vérité, et des conditions de vérité des phrases d'un langage, au sein d'une théorie de la signification, alors que les seconds portent sur la théorie de l'interprétation et ses conséquences[1]. J'ai, pour l'essentiel,

1. Davidson se consacre néanmoins plus spécifiquement aux conditions formelles de (a) dans 1965, 1967, 1969, 1970, et 1973 et aux conditions empiriques dans 1973, 1974, 1975.

respecté cette division dans ce qui suit, en consacrant le présent chapitre aux relations entre théorie de la signification et théorie de la vérité, et le chapitre suivant à la théorie de l'interprétation.

1.2. Conditions constitutives d'une théorie de la signification

L'objectif principal d'une TS pour une langue naturelle est donc de nous fournir une représentation de la connaissance que les locuteurs ont de leur langue, et du savoir qu'ils utilisent quand ils interprètent les énoncés des autres locuteurs. Les principaux réquisits d'une théorie de la signification découleront des faits qu'on peut tenir comme les plus généraux de la capacité à comprendre et à interpréter un langage. Davidson dégage d'abord quatre principes.

1 / Un locuteur compétent d'un langage *L* est normalement en mesure d'interpréter toutes les phrases de son langage, c'est-à-dire de leur assigner une signification. Si une TS pour *L* doit représenter cette compétence, elle devra également assigner à chaque phrase de L sa signification, *i.e* nous « fournir une méthode pour décider, étant donné une phrase arbitraire, quelle est sa signification » (1970 : 56, *95*). C'est ce que Davidson appelle la condition de *scrutabilité* d'une théorie de la signification. Elle revient à nous assurer que nous avons bien assigné aux expressions de *L* leurs significations sur la base de la façon dont elles sont désignées dans la théorie.

2 / Comprendre une expression linguistique, c'est être capable de discerner sa structure, c'est-à-dire la manière dont sa signification est déterminée par celle de ses parties composantes. Une TS doit donc être compositionnelle, ou satisfaire le principe sémantique de *compositionnalité*.

3 / Cette condition est étroitement associée à une autre : un locuteur compétent de L est normalement en mesure de comprendre, à partir d'un ensemble fini d'expressions, un ensemble infini d'autres expressions complexes et composées à partir de ces expressions (1967 : 17, *41*). Comprendre un langage, en ce sens, c'est être capable de l'apprendre.

Un langage ne peut être compris ni appris si l'on ne suppose pas que, d'une manière ou d'une autre, la signification de ses expressions complexes est fonction de leur structure et de l'application d'un certain nombre d'opérations sémantiques à un certain nombre d'éléments sémantiques. Il est essentiel que ces opérations et éléments soient en nombre fini pour qu'un langage puisse être appris. Appelons cette condition celle de *finitude*.

4 / Une quatrième condition, qui découle de la démarche de Davidson, 1967, et de la stratégie indirecte qui a été énoncée ci-dessus, est qu'une TS doit recourir à un certain concept, désigné comme central, en termes duquel cette théorie doit être formulée. De quoi en effet une telle théorie est-elle la théorie, et qu'est-ce qui est compris quand on comprend un langage ? La réponse semble triviale : le concept central est celui de signification, ce qui est compris, ce sont des significations. Mais cette réponse est également circulaire : que pouvons-nous apprendre d'une TS, si ce qu'on nous dit est qu'elle est une théorie des significations ? Comme on l'a vu, une TS devra être formulée de manière à éviter l'emploi, dans le langage utilisé pour formuler la théorie, des concepts mêmes qui doivent être expliqués. On peut appeler cette condition condition d'*immanence*[1]. En particulier on ne pourra pas présupposer le concept de signification lui-même. Il y a donc tout lieu de penser qu'une TS devra recourir à un autre concept désigné comme central, auquel s'appliqueront les autres conditions, comme celle de compositionnalité. Ce concept devra avoir un lien étroit avec celui de signification, c'est-à-dire être un concept sémantique, susceptible de jouer le rôle de valeur sémantique des expressions.

Ces quatre conditions forment la base de ce qu'on doit attendre d'une théorie de la signification. A l'exception de celle d'immanence, elles sont admises par un grand nombre de théoriciens du langage et de la signification, et, sous cette forme encore générale, elles ne suffisent pas à caractériser ce qu'il y a de spécifique à la conception de Davidson. Il ajoute trois autres conditions, beaucoup plus sujettes à controverse.

1. J'emprunte ce terme à Laurier, 1983.

5 / Le concept central auquel devra recourir une TS est celui de *vérité*. Connaître la signification d'une phrase, c'est au moins connaître les conditions dans lesquelles cette phrase est vraie ou fausse, c'est-à-dire ses conditions de vérité. On dira en ce sens qu'une TS doit être *véricon-ditionnelle*. Il y a cependant deux versions distinctes de cette thèse. Selon ce que l'on peut appeler sa version *forte*, la signification d'une phrase est constituée par ses conditions de vérité, qui déterminent cette signification, ou auxquelles elle s'identifie : *s* signifie que *p* si et seulement si *p* est vraie si c'est le cas que *p* (les conditions de vérité de *s* sont les conditions nécessaires et suffisantes de la signification de *s*). Selon la version *faible* de la véricondi-tionnalité, la signification d'une phrase détermine ses conditions de vérité, sans que celles-ci la déter-minent : si *s* signifie que *p*, alors *s* est vraie si c'est le cas que *p* (les conditions de vérité de *s* sont les conditions nécessaires de la significa-tion de *p*). On laissera pour le moment indéterminée la question de savoir laquelle des deux versions est défendue par Davidson. Le point important est ici seulement qu'une TS devra, d'une manière ou d'une autre, reposer sur une théorie de la vérité, ou des conditions de vérité.

6 / Comprendre une expression, c'est d'après le principe de compo-sitionnalité, comprendre les expressions plus simples dont elle est compo-sée. Mais il y a deux lectures possibles de ce principe. On peut le lire comme impliquant que la signification d'une phrase est déterminée par celle de ses parties composantes, mais non pas inversement. En ce sens, on soutient une version *atomiste* du principe. Mais on peut le lire aussi comme impliquant que la signification des parties composantes d'une phrase est déterminée elle-même par la signification de la phrase entière. En ce sens, on soutient une version *holiste* du principe de compositionnalité. Quelle que le soit le sens du fameux principe « contextuel » de Frege, selon lequel « ce n'est que dans le contexte d'une phrase que les mots ont une signification »[1], c'est dans ce sens holistique que Davidson le comprend. On peut ici parler d'un *holisme de la phrase*. A ce holisme s'en ajoute un

1. Frege, 1883, trad. fr. 122. Le principe fregéen est lui-même sujet à nombre d'interpréta-tions. Cf. Dummett, 1973 ; Dummett, 1982.

autre : comprendre un langage, c'est comprendre l'ensemble des significations des phrases du langage, et l'ensemble des éléments qui les composent. Une phrase n'a jamais de signification indépendamment du contexte de l'ensemble des autres phrases d'un langage, et les expressions qui les composent n'acquièrent de sens que dans le contexte de toutes les phrases où elles figurent. On peut appeler ceci *holisme du langage*. Le passage effectué par Davidson du premier au second holisme est explicite :

> Si la signification des phrases dépend de leur structure, et si nous ne compre-nons la signification de chaque élément de la structure que par abstraction à partir de la totalité des phrases dans lesquelles il figure, alors nous ne pouvons donner la signification d'une phrase quelconque (ou d'un mot) qu'en donnant la signification de toutes les phrases (et mots) du langage. Frege disait que ce n'est que dans le contexte d'une phrase qu'un mot a un sens ; dans la même veine, il aurait pu ajouter que ce n'est que dans le contexte du lan-gage qu'une phrase (et par conséquent un mot) a une signification (1967 : 22, *48*).

7 / Si comprendre le sens d'une expression, c'est comprendre le rôle qu'elle joue et la contribution qu'elle apporte au sens d'une phrase, et si com-prendre une phrase c'est comprendre le rôle qu'elle joue au sein d'un ensem-ble d'autres phrases, de quelle nature peuvent être ces rôles ou ces contributions ? Les seuls termes vagues de « composition » ou de « structure » employés jusqu'ici sont insuffisants. Il faut dire encore de quelle sorte de composition ou de structure il s'agit. La suggestion la plus évidente est que les phrases ont une certaine structure grammaticale, en vertu de laquelle elles sont articulées. Une TS devra donc nous per-mettre de discerner la structure grammaticale des phrases, et la récurrence de cette structure au sein de l'ensemble des phrases. Mais cela ne suffit pas à rendre compte des relations que, en vertu de l'hypothèse holistique, les phrases ont entre elles. De quelles sortes de relations peut-il s'agir ? Davidson suggère que ces relations sont, avant tout, des relations logi-ques ou inférentielles : de certaines phrases on peut en inférer d'autres, et ces relations logiques ont un rôle important à jouer dans la signification des phrases. Mais ces relations logiques ne peuvent elles-mêmes exister qu'en vertu du caractère des éléments qui composent ces phrases. Cela

suggère que la structure compositionnelle devra aussi avoir certaines propriétés logiques, et qu'elle s'identifiera, au moins jusqu'à un certain point, avec la structure logique ou la forme logique discernable dans ces phrases. En ce sens, déterminer la composition des phrases, c'est révéler leur forme logique, et le pouvoir inférentiel qu'ont les phrases et les expressions :

> Je voudrais donner une analyse du rôle logique ou grammatical des parties ou des mots [des] phrases qui s'accorde avec les relations d'implication entre [les] phrases et avec ce que l'on sait du rôle de ces mêmes parties ou mots dans d'autres... phrases. Je tiens cette entreprise comme étant identique à celle consistant à montrer comment la signification des phrases... dépend de leur structure (1967a, 106, *149-150*).

Comprendre un langage, c'est connaître sa « géographie logique », et

> donner la forme logique d'une phrase, c'est donner sa localisation logique au sein de la totalité des phrases, la décrire d'une manière qui détermine explicitement quelles phrases elle implique et par quelles phrases elle est impliquée (1967a, 140, *190*).

Ces sept conditions, de scrutabilité, de compositionnalité, de finitude, d'immanence, de vériconditionnalité, de holisme et de forme logique, figurent toutes, de manière plus ou moins explicite, dans les principaux articles où Davidson présente son programme. Comme on l'a vu, elles sont étroitement liées entre elles, et elles dessinent la forme générale d'une théorie de la signification. On n'en aura une idée plus précise que si l'on suit l'argumentation de « Truth and Meaning ».

1.3. Conditions formelles d'une théorie de la signification

Nous cherchons à déterminer la forme d'une théorie capable d'assigner une signification à toutes les phrases d'un langage donné. Une théorie est elle-même formulée dans un certain langage, que nous pouvons désigner, en un sens encore vague, comme « métalangage » par rapport au langage dont elle est la théorie, le « langage-objet ». Mais nous ne

savons pas quelle est « la forme caractéristique » des attributions de signification. Par exemple, ces attributions doivent-elles être directes, et avoir la forme « La signification de l'expression... est... », ou la forme « L'expression... signifie... », ou bien ces attributions doivent-elles être indirectes et établir les significations d'expressions par le biais d'autres expressions qui leur soient synonymes ou qui en soient la traduction ? D'une manière ou d'une autre, notre théorie doit inclure un moyen de désigner des expressions. Cette condition est remplie si les expressions du langage-objet sont désignées par des « descriptions structurales », c'est-à-dire des descriptions construites à partir des noms des symboles du langage-objet unies par un signe de concaténation (Tarski, 1956). Mais une fois que nous disposons de ce mode de désignation des expressions du langage-objet, comment allons-nous établir leurs significations ? Devrons-nous en particulier construire nos attributions de signification de manière à désigner également non pas des expressions, mais des significations ? Davidson envisage deux suggestions courantes à cet effet.

La première consiste à donner aux attributions de signification la forme :

(i) *s* signifie *m*

où « *s* » est une description structurale d'une phrase du langage-objet, et « *m* » un terme singulier faisant référence à la signification de *s*. Nous satisfaisons ainsi apparemment à la condition de scrutabilité, puisque cela devrait nous donner une « méthode effective » pour parvenir à la signification d'une phrase quelconque. Si « *m* » est un terme singulier désignant une signification, cette dernière est une certaine entité. Davidson qualifie de « fregéenne » la conception selon laquelle les significations seraient des entités. Telle qu'il la comprend, elle consiste à discerner la structure des phrases comme comportant des expressions de fonctions (insaturées) et des expressions venant occuper la place d'argument, puis assigner à chaque partie d'une phrase une entité qui est sa signification. Mais cette solution ne nous donne aucun moyen d'expliquer la compositionnalité des significations :

> On se demande, par exemple, quelle est la signification de « Théétète vole ». Une réponse fregéenne pourrait être la suivante : étant donné la signification de « Théétète » comme argument, la signification de « vole » donne la signification de « Théétète vole » comme valeur. La vacuité de la réponse est évidente. Nous voulions savoir ce qu'est la signification de « Théétète vole » ; cela ne nous avance à rien d'apprendre que c'est la signification de « Théétète vole »... Paradoxalement, la seule chose que les significations ne fassent pas est d'huiler les roues d'une théorie de la signification — tout au moins tant que nous requérons d'une telle théorie qu'elle nous donne non trivialement la signification de chaque phrase dans le langage. Mon objection aux significations dans une théorie de la signification n'est pas qu'elles soient abstraites ou que leurs conditions d'identité soient obscures, mais qu'elles n'aient pas d'usage démontré (1967 : 20, *46*).

Davidson entend ainsi dissocier son objection de celle de Quine contre les significations comme entités intensionnelles[1]. Ce que nous devons attendre selon lui d'une théorie de la signification est qu'elle nous « donne la signification » des phrases d'un langage en montrant comment est déterminée leur structure compositionnelle. Cela ne peut manifestement pas être réalisé si l'on se contente d'assigner à chaque partie de phrase une signification comme entité. Seule une analyse qui établirait comment la signification des phrases est déterminée compositionnellement à partir des éléments qui affectent leurs conditions de vérité peut « effectivement » « donner » leur signification[2].

La seconde suggestion consisterait à donner aux attributions canoniques de signification la forme d'une traduction des phrases du langage-objet dans le métalangage de la théorie. Ces attributions auraient ainsi la forme suivante :

(ii) « *p* » (dans *L'*) traduit « *s* » (dans *L*)

En d'autres termes, on propose qu'une théorie de la signification prenne la forme d'un manuel de traduction de *L* dans une autre langue *L'*. *A priori* rien n'interdit à un tel manuel de traduction de nous permettre

1. Cf. par exemple, Quine, 1960 : chap. 6 ; Quine, 1969.

2. Notons au passage qu'il est curieux, de ce point de vue, d'opposer à une telle analyse la conception « fregéenne » visée ici par Davidson, puisque Frege défend lui-même une analyse à la fois vérifonctionnelle et compositionnelle de la signification. Cf. Engel, 1985, chap. 2.

d'interpréter et de comprendre des phrases de *L*, puisque quiconque comprend *L'* devrait être en mesure, si la traduction est correcte, de comprendre *L*. Mais la difficulté ne tient pas au fait que nous pouvons nous demander quels sont les critères d'un bon manuel de traduction, ni même que l'emploi d'un tel manuel reviendrait à présupposer la notion de synonymie et de signification. La difficulté tient au fait que la possession d'un manuel de traduction d'une langue dans une autre ne nous permettrait pas de savoir ce que les phrases de *L* signifient si nous ne savions pas ce que signifient celles de *L'*. En d'autres termes, quelqu'un qui saurait que

(i) « Kiler ne radi kako treba » signifie (traduit) en serbe ce que « The radiator does not work » signifie en anglais

n'aurait aucune idée de ce que signifie la phrase serbe en question s'il ne savait pas ce que signifie la phrase anglaise. Cette difficulté affecte une certaine conception des théories sémantiques qui a été avancée dans le cadre de la linguistique transformationnelle. Katz et Postal (1964) ont proposé qu'une théorie sémantique pourrait consister, sur la base d'une grammaire transformationnelle spécifiant la syntaxe d'une langue, en un ensemble de traductions des expressions de cette langue dans un langage fournissant pour chaque entité lexicale spécifiée son « marqueur sémantique » en vertu de définitions « analytiques ». Ici encore, le problème ne tient pas à la notion d'analyticité qui sert à encoder les expressions syntaxiquement spécifiées par des définitions de dictionnaire dans le langage « marqueur-sémantiquais ». Il tient au fait que, comme dans le cas du manuel de traduction, on pourrait disposer des bonnes règles d'association des entités syntaxiques aux significations spécifiées en marqueur-sémantiquais sans pour autant savoir ce que les phrases signifient (1967 : 21, *47*). Davidson illustre ce point avec le cas des phrases rapportant des attitudes propositionnelles comme la croyance. La connaissance de la signification lexicale du verbe « croire » ne suffit pas pour comprendre la signification d'expressions de la forme « X croit que *p* », car elle ne nous dit en rien quelles sont les conditions de vérité et la forme logique de telles expressions. Ce qui est visé ici n'est pas seulement toute conception qui ferait d'une théorie sémantique simplement un ajout à une théorie de la structure syntaxique des phrases, mais également toute théorie

sémantique qui prétendrait se passer de notions comme celle de conditions de vérité. David Lewis adresse sur ce point la même critique aux sémantiques traductionnelles :

> Nous pouvons savoir quelle est la traduction d'une phrase du français en marqueurais sans avoir la moindre idée de la signification de la phrase française : à savoir les conditions sous lesquelles elle serait vraie. Une sémantique sans conditions de vérité n'est pas une sémantique (Lewis, 1972 : 169).

Nous ne devons cependant pas en conclure qu'une théorie de la signification pour un langage *L* ne nous fournit pas une traduction, dans le métalangage de la théorie, des phrases de *L*. Rappelons que Davidson exige d'une théorie de la signification que sa connaissance puisse suffire pour interpréter et comprendre un langage. Il n'y a rien à objecter, de ce point de vue, à une théorie qui nous donnerait des spécifications du type :

(2) « Kiler ne radi kako treba » signifie (en serbe) que le radiateur ne marche pas.

La différence entre (1) et (2) est que la phrase française qui nous donne la signification de la phrase serbe mentionnée n'est pas elle-même mentionnée ; elle est utilisée. Nous la comprenons parce que nous comprenons le français, et parce que la relation entre la phrase française et la phrase serbe est une relation de traduction ou de synonymie, mais ce n'est pas pour autant une théorie de la traduction. Dans (1) au contraire, la phrase anglaise qui traduit la phrase serbe est mentionnée. Cela montre qu'une théorie de la signification pour *L* formulée dans un second langage n'est pas une théorie de la traduction. Elle implique que nous comprenions le langage dans lequel elle est formulée. Il serait absurde de supposer qu'une théorie de la signification ne puisse pas être articulée dans un langage quelconque, et que ce langage ne doive pas être compris par ceux qui connaissent cette théorie. Le problème n'est donc pas de chercher à éviter tout recours à une notion comme celle de traduction, mais de spécifier les règles précises qui nous permettraient d'obtenir une relation qui ait le même effet que la notion de traduction.

Puisque des théories de la signification utilisant les attributions canoniques (i) et (ii) se révèlent inadéquates, pourquoi ne pas retenir simplement les attributions de la forme de (2), c'est-à-dire :

(iii) « *s* » signifie (dans *L*) que *p* ?

Cette démarche est parfaitement légitime. Elle implique que nous énoncions à quelles conditions la phrase de *L* mentionnée à gauche de « signifie que » a la même signification que la phrase du métalangage employée à droite. Les tentatives en ce sens ne manquent pas. L'une des plus caractéristiques est celle de Carnap dans *Meaning and Necessity* (1956), qui s'efforce de définir l'équivalence en signification entre deux expressions. Selon Carnap, cette équivalence ne peut pas être une équivalence extensionnelle, au sens où la phrase de gauche aurait la même valeur de vérité que la phrase de droite. Il est évident en ce sens que de

(3) « La neige est blanche » signifie que la neige est blanche

et de

(4) « La neige est blanche » et « L'herbe est verte » ont la même valeur de vérité

on ne peut inférer que

(5) « La neige est blanche » signifie que l'herbe est verte.

Carnap cherche à résoudre ce problème en proposant une définition distincte de l'équivalence, la L-équivalence. Le problème n'est pas ici celui discuté par les critiques de Carnap, comme Quine (1953), de savoir si cette définition est adéquate et non circulaire. Il tient seulement à ceci que, comme le dit Davidson, si nous cherchons à « batailler avec la logique du terme apparemment non extensionnel "signifie que" nous rencontrerons des problèmes aussi difficiles que ceux que notre théorie cherche à résoudre » (1967 : 22, *48*). C'est ici qu'intervient la condition d'immanence, qui nous prescrit de formuler une théorie de la signification sans recourir explicitement au concept de signification et par conséquent de substituer un autre prédicat sémantique au « signifie que » de (iii), sans pour autant renoncer à l'idée que la phrase de droite du métalangage établisse la signification de la phrase de gauche du langage-objet :

> La théorie aura fait son travail si elle fournit, pour chaque phrase du langage étudié, une phrase correspondante (remplaçant « *p* ») qui, d'une certaine

manière qui doit encore être clarifiée, « donne la signification » de *s* (1967 : 23, *49*).

Deux candidats se présentent naturellement. La phrase de droite peut être *s* elle-même, comme dans (3), auquel cas la théorie de la signification est ***homophonique***, c'est-à-dire établit la signification de la phrase de gauche dans le même langage que cette phrase. La phrase de droite peut être aussi une traduction de la phrase de gauche dans un autre langage, comme dans (2), auquel cas notre théorie de la signification est dite ***hétérophonique***. Mais nous devons éliminer le « signifie que ». La solution préconisée par Davidson est alors :

> Sautons finalement le pas audacieusement, et essayons de traiter la position occupée par « *p* » extensionnellement : pour réaliser cela, écartez l'obscur « signifie que », donnez à la phrase qui remplace « *p* » un connecteur propositionnel approprié, et donnez à la description que remplace « *p* » son propre prédicat. Le résultat plausible est
>
> (*T*) *s* est *T* si et seulement si *p*.
>
> Ce que nous demandons à une théorie de la signification pour un langage L est que sans faire appel à aucune notion sémantique (supplémentaire) elle place suffisamment de restrictions sur le prédicat « est *T* » pour impliquer toutes les phrases obtenues à partir du schéma T quand « *s* » est remplacée par une description structurale d'une phrase de L et « *p* » par cette phrase.
>
> Deux prédicats satisfaisant cette condition ont la même extension, en sorte que si le métalangage est suffisamment riche, rien ne s'oppose à ce que nous donnions « est *T* ». Mais qu'on le définisse explicitement ou qu'on le caractérise récursivement, il est clair que les phrases auxquelles le prédicat « est *T* » s'applique seront justement les phrases vraies de L, parce que la condition que j'ai imposée aux théories satisfaisantes de la signification est dans son essence la Convention *T* de Tarski qui teste l'adéquation d'une définition formelle de la vérité (1967 : 23, *49-50*).

La condition formelle principale d'adéquation d'une TS est précisément la Convention *T*, qui est pour Tarski la « condition d'adéquation formelle » d'une théorie de la vérité, et une TS doit devenir, en un sens qu'il faudra préciser, une théorie de la vérité (ce que l'on appellera désormais une « théorie-T »). Ce passage énonce, mais ne montre pas qu'une TS doit prendre la forme d'une théorie de la vérité au sens de Tarski. Davidson ne montre pas qu'une TS doit être extensionnelle. Il n'établit pas que le connecteur « si et seulement si » est le connecteur extensionnel approprié. Il ne montre pas non plus que le prédicat de vérité est le seul candidat approprié pour prendre la place de « est T »[1]. S'il y a un argument ici, il n'est pas déductif ; c'est plutôt ce que l'on appelle une « inférence vers la meilleure explication », qui nous suggère de traiter les réquisits de Tarski pour une définition de la vérité comme ceux qui permettent de satisfaire le mieux les conditions constitutives de Davidson : en particulier être scrutable, immanente, compositionnelle et vériconditionnelle. Mais soutenir qu'une TS peut gagner à être mise sous la forme d'une théorie-T au sens de Tarski, et soutenir qu'une TS *est*, ou s'identifie purement et simplement avec une théorie-T au sens de Tarski, sont deux choses différentes. Davidson ne soutient pas la seconde thèse, mais seulement la première : *dans certaines conditions*, une théorie-T pour *L* peut *jouer le rôle* d'une TS pour *L*.

Il y a un cas où cette exigence est clairement remplie : une TS pour un langage L est bien une théorie-T, et sous des conditions tout à fait spécifiables, lorsque L est un langage *formel*. C'est précisément ce que Tarski a cherché à établir. En posant le problème pour le cas où *L* est une langue ***naturelle***, Davidson change sa nature, mais du même coup sa proposition paraît infondée, et non pas simplement « audacieuse », si elle revient simplement à transposer aux langues naturelles les conditions de Tarski pour la construction d'une sémantique pour les langues formelles. Davidson révise donc ces conditions.

Davidson s'écarte de Tarski sur quatre points : (a) le rejet de l'idée qu'une langue naturelle est susceptible d'une définition formellement correcte, (b) le rejet de la notion de traduction dans la formulation de la Convention *T*, (c) le rejet de l'idée d'une définition explicite de la vérité au bénéfice de la seule notion de théorie de la vérité, et (d) l'adaptation de la convention *T* aux expressions indexicales.

(a) Dans son *Wahrheitsbegriff*, Tarski entreprend de donner une définition du terme « phrase vraie » dans un langage, une définition qui

1. Comme le remarque Platts, 1980 : 55-56.

soit « matériellement adéquate » et « formellement correcte »[1]. Une définition de la vérité est formulée dans un métalangage, pour un langage-objet. Elle ne peut être formulée que pour un langage dont la structure formelle soit déjà spécifiée. Ces deux traits excluent qu'on puisse fournir une définition de « vrai » pour une langue naturelle : les langues naturelles n'ont pas de structure formelle spécifiée, et elles sont « universelles », c'est-à-dire autorisent, à l'intérieur même de ces langues, la référence à des expressions de ces langues, ce qui conduit à des antinomies comme celle du Menteur. Seule une langue formelle peut recevoir une définition correcte de la vérité. A cela, Davidson répond que si l'on ne voit pas en quoi les ressources expressives d'une langue naturelle pourraient nous permettre de donner une définition explicite de la vérité dans cette langue, on ne voit pas non plus pourquoi les langues naturelles devraient servir à donner ce genre de définitions (1967 : 29, *58-59*). En d'autres termes, il est peut être possible de caractériser, pour une langue naturelle, quelque chose comme la vérité sans pour autant que cette caractérisation prenne la forme d'une *définition* de la vérité. De plus, comme le souligne Davidson, son projet ne vise pas à donner une définition de la vérité pour une langue naturelle prise dans son hypothétique totalité, mais des théories pour divers fragments d'une telle langue, suffisants pour illustrer les relations qu'il entend établir entre une théorie de la vérité et une théorie de la signification. Cela revient bien à admettre que la nature du problème posé par la notion de vérité appliquée à une langue naturelle et à une langue formelle respectivement n'est pas le même selon qu'il s'agit de l'une ou l'autre. Cela ne veut pas dire qu'il n'y ait pas des similarités importantes entre les deux cas, ni que l'on ne puisse adapter les réquisits formels qui valent pour les langues formelles aux langues naturelles. Tarski ignore systématiquement la possibilité de cette adaptation.

A l'objection selon laquelle les langues naturelles n'ont pas une structure formelle spécifiée, Davidson répond que l'essentiel n'est pas de réformer la langue naturelle au moyen d'une langue canonique meilleure que « l'idiome brut » mais de le comprendre. Exiger qu'une théorie de la

1. Dans tout ce qui suit, je présuppose de la part du lecteur une connaissance de la structure globale d'une théorie de la vérité selon Tarski. Pour une introduction, cf. Quine, 1970 ; Engel, 1989, chap. 5 (qui contient des erreurs, cf. de préférence l'édition de 1991).

signification prenne la forme d'une théorie-T ne présuppose pas que le langage-objet ait une structure formelle spécifiée, ni que le métalangage de la théorie soit lui-même un langage canonique. « Si nous savons *pour quel* idiome la notation canonique est canonique, nous avons une aussi bonne théorie pour l'idiome que pour son compagnon mis sous tutelle » (1967 : 29, *58*).

Reprenons l'exemple de la phrase (2) ci-dessus, adaptée de manière à être conforme à la Convention T :

(2') « Kiler ne radi kako treba » est vrai (en serbe) ssi le radiateur ne marche pas.

Ici notre métalangage est le français. Si nous avons une façon quelconque d'établir que les conditions de vérité de la phrase de gauche sont les mêmes que celles de la phrase de droite, alors nous avons satisfait à nos réquisits. Supposons maintenant que nous ayons affaire à une phrase française à gauche, et que la phrase de droite du métalangage soit formulée dans la notation canonique du calcul des prédicats du premier ordre :

(6) « Il y a des choux et il y a des rois » est vrai ssi (∃ x) (x est un chou) & (∃ y) (y est un roi)

Dans les conditions imposées par Davidson, rien ne nous engage à choisir cette notation canonique, bien qu'il y ait des raisons importantes de penser qu'une notation canonique révèle les conditions de vérité. Mais en aucun cas il n'est requis que cette notation canonique soit considérée comme une traduction, dans un langage idéal, de la phrase de la langue correspondante. Ce qui importe est que les conditions de vérité de l'une et celles de l'autre s'accordent, et qu'on ait un moyen d'établir cette correspondance. Nous devons donc renoncer à la condition de Tarski selon laquelle la vérité doit, pour une langue naturelle, recevoir une définition formellement correcte.

(b) En second lieu, la Convention *T* est chez Tarski une condition d'adéquation matérielle d'une définition de la vérité. Elle requiert qu'une définition de « phrase vraie » (pour *L*) ait comme conséquences des phrases de la forme :

(*T*) *S* est vrai (dans *L*) ssi *p*

qui seront les théorèmes d'une théorie de la vérité. Appelons désormais celles-ci des « *phrases-T* ». Selon Tarski, la Convention T implique que la phrase « *p* » du métalangage située à droite du biconditionnel soit une *traduction* de la phrase « *s* » du langage-objet située à gauche. Davidson ne requiert rien de tel. Présupposer la notion de traduction, ou celle d'identité de signification, ce serait présupposer la notion de signification, et violer la condition d'immanence. Davidson ne requiert qu'une chose : que les phrases de gauche et celles de droite aient la même valeur de vérité, en sorte que le biconditionnel correspondant soit vrai :

> Dans l'œuvre de Tarski, les phrases-T sont tenues pour vraies parce que le côté droit du biconditionnel est supposé être une traduction de la phrase dont les conditions de vérité sont établies... Ce que je propose est de renverser l'ordre d'explication : en supposant la traduction donnée, Tarski était capable de définir la vérité ; l'idée est de traiter la vérité comme primitive et d'extraire une analyse de la traduction ou de l'interprétation...
>
> Il n'y a pas de difficulté à reformuler la Convention *T* sans faire appel au concept de traduction : une théorie acceptable de la vérité doit impliquer, pour chaque phrase s du langage-objet, une phrase la forme : *s* est vraie si et seulement si *p*, où « *p* » est remplacée par toute phrase qui est vraie si et seulement si *s* l'est (1973 : 134, *199*).

(c) Tarski entend donner une définition explicite de la vérité pour un langage formel, c'est-à-dire éliminer le concept de vérité et les autres concepts sémantiques utilisés dans les premières étapes de sa construction, de manière à ce que ni « est vrai », ni d'autres concepts comme « désigne » ou « a comme référence » n'apparaissent à gauche des biconditionnels des phrases-T. Une définition récursive de la vérité est donnée sur la base du concept de satisfaction. Mais une définition explicite ne peut être établie que si les ressources du métalangage sont étendues fortement au-delà de celles du langage-objet. Chez Davidson, la Convention *T* teste l'adéquation non pas d'une *définition*, mais d'une *théorie* de la vérité. Comme on l'a vu, le prédicat de vérité reste non défini, et il peut survenir à gauche des biconditionnels de la forme (*T*). Ainsi il n'est pas question d'éliminer les notions sémantiques primitives, comme celles de satisfaction et de vérité. On doit donc se restreindre à une classe limitée de théories de la vérité, celles dont les concepts ou l'« idéologie » n'excèdent pas en ressources expressives celles du métalangage (1973a : 72, *116-117*).

La Convention *T* n'est donc pas un moyen d'éliminer le prédicat de vérité. Elle n'a donc pas non plus, pour Davidson, les conséquences philosophiques que certains philosophes entendent tirer de la « théorie sémantique » de la vérité de Tarski (cf. *infra*, chapitre 5). C'est à une caractérisation, purement extensionnelle, du prédicat « vrai » que servent les théories-T :

> Les phrases-T ne montrent pas comment vivre sans prédicat de vérité ; mais une fois réunies elles disent ce à quoi cela ressemblerait d'en avoir un. Car puisqu'il existe une phrase-T qui correspond à chaque phrase du langage pour lequel la vérité est en question, la totalité des phrases-T fixe l'extension, parmi les phrases, de tout prédicat qui joue le rôle des mots « est vrai ». il est donc clair que même si les phrases-T ne définissent pas la vérité, elles peuvent être utilisées pour définir ce que c'est qu'être un prédicat de vérité ; tout prédicat qui rend toutes phrases-T vraies est un prédicat de vérité (1973a : 65, *108*).

(d) Davidson fait subir à la Convention *T* une quatrième modification. L'une des différences les plus évidentes entre les langues formelles et les langues naturelles est que ces dernières contiennent des expressions indexicales. Les conditions de vérité et de référence des phrases qui contiennent ces expressions ne peuvent être établies que si l'on tient compte de l'identité des locuteurs, des circonstances et du temps de l'énonciation. Davidson propose qu'à cet effet on relativise le prédicat « est vrai » à un locuteur et à un temps donné. « Vrai » devient alors un prédicat à trois places, reliant une phrase, une personne et un temps. La théorie impliquera alors des phrases comme

> (7) « Je suis fatigué » est vrai en tant que (potentiellement) énoncé par P en *t* si et seulement si P est fatigué en *t* (1967 : 34, *66*).

Le problème n'est pas ici de savoir si cette proposition a des chances de résoudre le problème qu'elle est supposée résoudre. Mais il est clair qu'elle est de nature à changer sérieusement les réquisits de Tarski.

Il est par conséquent difficile de dire que les conditions de Davidson reviennent à appliquer purement et simplement les critères formels d'une théorie-T tarskienne à des langues naturelles. Les théories-T en question différeront grandement selon que l'on aura affaire à une langue formelle ou à une langue naturelle. Quand Davidson soutient hardiment que « si

nous réussissons à donner une théorie formelle de la vérité pour une langue naturelle, nous considérerons le langage naturel comme un langage formel », il faut sérieusement relativiser ces déclarations, comme il le fait d'ailleurs lui-même[1].

Finitude. — On commentera à présent les autres conditions formelles. L'une des conditions constitutives est celle de *finitude*. Sa traduction formelle est que les théories-T qui sont supposées jouer le rôle de TS soient des théories pourvues d'un nombre fini d'axiomes. Cette condition ne figure pas chez Tarski, puisqu'il définit la vérité explicitement dans un métalangage qui ne contient pas d'axiomes en nombre fini pour des expressions primitives. En pratique la différence est minime, puisqu'il existe une procédure, due à Frege et Dedekind, pour remplacer une définition récursive en une définition explicite. Comme on l'a vu, Davidson justifie avant tout cette condition par la nécessité de permettre l'apprentissage des langues dont elles sont les théories. La nature exacte de l'argument de Davidson pour relier la condition d'apprentissage à la notion d'une théorie-T avec un nombre fini d'axiomes n'est pas totalement claire. Tantôt l'argument semble reposer sur une considération empirique et psychologique portant sur la nature de nos capacités cognitives : comment des créatures finies comme nous le sommes, capables d'emmaganiser seulement une quantité finie d'informations peuvent-elles comprendre une infinité potentielle de sens nouveaux ? Tantôt l'argument ressemble plus à une analyse des conditions *a priori* d'un apprentissage : un langage caractérisé par une théorie finie est —qu'il soit en fait appris ou non — susceptible d'être appris. Ce point est distinct du précédent, parce que des créatures qui auraient des capacités infinies de stockage d'information mais qui n'auraient pas d'autres capacités infinies ne pourraient pas acquérir beaucoup de connaissances finies en un temps fini[2]. L'argument est encore différent quand il a recours à la condition de forme logique :

> Que devrions-nous exiger d'une analyse adéquate de la forme logique d'une phrase ? Avant tout, à mon sens, une telle analyse devrait nous conduire à voir

1. Cf. par exemple, 1970 : 55, 99.
2. Comme le remarque Wright, 1986 : 206.

PATCH/UFMG
BIBLIOTECA

> le caractère sémantique de la phrase — sa vérité ou sa fausseté — comme dû au fait qu'il est composé par un nombre fini de quelques-uns parmi un nombre fini de dispositifs qui suffisent pour le langage dans son entier, à partir d'éléments tirés d'un stock fini (le vocabulaire) qui est suffisant pour le langage dans son ensemble. Considérer une phrase de cette manière, c'est la considérer à la lumière d'une théorie pour son langage, une théorie qui donne la forme de chaque phrase dans le langage. Une manière de fournir une telle théorie est de caractériser récursivement le prédicat de vérité, selon les grandes lignes de la procédure proposée par Tarski (1969 : 94, 145).

Davidson donne plusieurs exemples de théories sémantiques qui auraient comme conséquence que le langage ou le fragment de langage qu'elles caractérisent ne peut être appris. Dans (1965), il en donne quatre : la citation, le discours indirect, les phrases rapportant des croyances, et la logique du sens et de la dénotation selon Church. On ne considérera ici que les deux premiers.

Il y a une infinité de citations possibles mises entre des guillemets que nous pouvons produire et comprendre. Selon une conception avancée par Quine (1940) et par Tarski (1956), des guillemets transforment une expression utilisée en un nom, ou un terme singulier, de cette expression citée ou mentionnée. La fonction des lettres dans les mots nommés devient ainsi purement contingente, et elles peuvent être, comme Tarski le propose, remplacées par un nom structural de l'expression. Cela n'a par conséquent pas de sens de dire, comme Quine, que des guillemets nomment « ce qu'il y a à l'intérieur » des guillemets. Ainsi chaque guillemet devient une expression primitive du langage. Il s'ensuit qu'un langage contenant des guillemets devient impossible à apprendre. Le second exemple est celui de la théorie du discours indirect de Scheffler (1954), selon laquelle toutes les phrases rapportant des contenus au discours indirect établissent des relations entre un locuteur et des expressions concrètes, de la forme

X a énoncé une que-la-neige-est-blanche énonciation

où « que-la-neige-est-blanche » est un prédicat (*type*) unaire prenant comme argument des inscriptions (*tokens*). Comme dans le cas des citations, les inscriptions ont une structure, mais on ne voit pas en quoi leur signification dépend de cette structure. Chaque nouveau prédicat est ainsi une expression

primitive du langage, qui ne peut être appris. Une théorie qui ne repose pas sur une axiomatisation finie ne remplirait pas la condition d'apprentissage, parce qu'elle engendrerait un ensemble infini d'expressions sémantiquement primitives, et surtout parce que les structures qu'elle utiliserait n'auraient pas de signification déterminée. Qu'est-ce qu'une expression sémantiquement primitive ? C'est une expression telle que « les règles qui donnent la signification des phrases dans lesquelles elle ne figure pas ne suffisent pas à déterminer la signification des phrases dans lesquelles elle figure » (1965 : 12, *31*). Mais il est difficile, à première vue, de comprendre exactement pourquoi une théorie devrait, pour être apprise, avoir des primitives sémantiques. N'y a-t-il pas des théories mathématiques qui n'ont pas d'axiomatisation finie, comme l'arithmétique de Peano ou la théorie des ensembles de Zermelo-Fraenkel, que nous pouvons néanmoins parfaitement apprendre ? Et pourquoi ne pourrait-on maîtriser des théories comme celle de la citation de Quine, ou comme celle de Scheffler ? Il est évident qu'au sens usuel du terme, de tels langages peuvent être appris[1]. Mais ce que Davidson a en vue est différent. Il soutient qu'il y a une relation intrinsèque entre comprendre une phrase et comprendre sa structure. Il suggère que des théories-T qui ne révéleraient pas, d'une manière ou d'une autre, une structure sémantique ne seraient pas véritablement des théories de notre compréhension du langage. En ce sens, la finitude des axiomes est un réquisit bien moins important que le fait que ces axiomes confèrent une certaine structure. Supposons, par exemple, qu'une théorie de la vérité (T1) pour un langage L_1 soit axiomatisée au moyen d'un schéma d'axiome unique :

(A) *s* est vrai (en français) si et seulement si *p*.

Cette théorie aurait, par définition, un ensemble fini d'axiomes. Elle aurait une infinité de phrases-T comme conséquences, et serait en ce sens conforme à ce que requiert Davidson. Mais nous supposons qu'un locuteur puisse connaître cette théorie. Or le schéma d'axiome (A) n'est pas une vérité que nous puissions connaître ; il a le même statut qu'une phrase ouverte. Nous pouvons connaître toutes les instances de (A) et savoir qu'elles sont

1. Cette objection est celle de Chihara, 1972. Cf. également Le Pore et Loewer, 1982.

vraies, mais pas savoir à quelles conditions elles le sont, parce que nous ne serions pas en mesure de déterminer quels éléments composent ces phrases. Il en est de même ici que dans le cas de l'emploi de la notion de traduction : tout comme nous pouvons savoir qu'une phrase est une traduction d'une autre phrase sans savoir ce que l'une ou l'autre de ces phrases signifient, nous pouvons savoir que toute instance de (A) est vraie sans savoir ce que veulent dire ces instances.

Le second exemple est emprunté à Gareth Evans (1981). Supposons que nous ayons affaire à un langage, L_2, de structure élémentaire. Il a un vocabulaire de dix noms : *a, b, c, ...j*, et de dix prédicats monadiques : *F, G, H,..., O*. On peut construire en tout 100 phrases possibles dans L_2, en combinant noms et prédicats. L_2 est donc fini. Considérons deux théories de la vérité possibles pour ce langage. Chacune d'elles sera finie, par définition. La première, T_2, prendrait la forme d'une simple liste des axiomes qui seraient constitués par toutes les phrases-T construites à partir des phrases de L_2 :

« F*a* » est vrai ssi Yul est chauve
« F*b* » est vrai ssi Kojack est chauve
...
« G*a* » est vrai ssi Yul est courageux
« G*b* » est vrai ssi Kojack est courageux

et ainsi de suite. T_2 traite chacune des phrases de L_2 comme non structurée, c'est-à-dire ne contient aucun axiome qui spécifie des primitifs sémantiques, comme des noms ou des prédicats, pour L_2... L'autre théorie, T_3, a 21 axiomes, un pour chaque « mot » du langage, et un axiome général, compositionnel, énonçant comment former une expression complexe à partir d'expressions simples. Il y a 10 axiomes de la forme :

« *a* » dénote Yul
« *b* » dénote Kojack

et 10 axiomes de la forme :

Un objet satisfait F ssi il est chauve
Un objet satisfait G ssi il est courageux

L'axiome compositionnel est

> Une phrase couplant un nom avec un prédicat est vraie ssi l'objet dénoté par le nom satisfait le prédicat.

On peut dériver de ces 21 axiomes un énoncé des conditions de vérité de chacune des cent phrases de L_2, c'est-à-dire les 100 phrases-T que T_2 prend comme axiomes. T_3 traite les phrases *Fa, Gb*, etc., comme structurées ; elle discerne dans ces phrases deux éléments distincts, le nom *a* ou *b*, etc., et le prédicat *F, G*, etc. Néanmoins T_2 et T_3 sont extensionnellement équivalentes : elles donnent la même valeur de vérité aux mêmes phrases. Elles ne sont cependant pas logiquement équivalentes, puisque les axiomes de T_3 ne peuvent pas être déduits des axiomes de T_2. En principe chacune des deux théories devrait satisfaire les réquisits de Davidson, puisqu'elles sont finies. On peut voir que si le langage L_2 contenait une infinité de phrases (par exemple si on l'étendait en ajoutant des connecteurs propositionnels), une axiomatisation de la forme de T_3 serait la bonne et exclurait T_2, puisqu'elle montrerait comment engendrer, à partir des structures simples présentes, une infinité de phrases. Mais même si Davidson a souvent tendance, comme Chomsky, à assimiler la créativité de la compréhension et de l'usage du langage avec le caractère infini de la production linguistique, l'exemple de T_2 et T_3 montre que ces deux traits ne sont pas essentiellement liés. Ce qui fait la supériorité de T_3 par rapport à T_2 est que cette dernière ne nous dit rien sur la structure du langage L_2 ; elle ne nous dit pas comment la signification d'une phrase est déterminée par celle de ses parties composantes. Il faut donc distinguer la condition de finitude des axiomes de ce que nous pouvons appeler la condition de structure sémantique[1]. Il est plus juste de dire que la première est une façon de satisfaire la seconde que l'inverse, et que c'est sur cette dernière que pèse le poids d'une analyse de la compréhension du langage.

Homophonie et hétérophonie. — Comme on l'a vu, la condition d'immanence prescrit de formuler une TS telle que les concepts à expliquer ne

1. J'emprunte ceci à Davies, 1981 : 58 ; cf. également Wright, 1987, chap. 6.

soient pas utilisés par la théorie. Une TS prenant la forme d'une théorie-T, la condition d'immanence signifie alors que les énoncés des conditions de vérité des phrases doivent reposer sur les mêmes concepts que ceux des phrases dont ils établissent les conditions de vérité. Cette caractéristique va de pair, selon Davidson, avec son interprétation de la Convention T comme critère d'une théorie, plutôt que d'une définition de la vérité, puisque le seul terme sémantique autorisé à droite des biconditionnels (T) est le prédicat de vérité lui-même, et qu'une définition demanderait des ressources expressives plus fortes du métalangage. Cela favorise systématiquement les théories *homophoniques*, c'est-à-dire les théories pour lesquelles le langage-objet est contenu dans le métalangage, ou pour lesquelles les ressources expressives du métalangage se rapprochent le plus possible de celles du langage-objet. Il existe une classe de théories sémantiques familières qui satisfont apparemment à la Convention T et à certaines des conditions de Davidson, mais qui ne sont pas homophoniques. Ce sont les théories de la vérité fondées sur la théorie des modèles, qui relativisent la vérité à une interprétation, à un modèle, ou à monde possible. Une théorie-T est dite absolue si la satisfaction des prédicats y est définie par des assignations de suites d'objets pris dans un domaine unique, quand elle produit des clauses comme

> Pour toute suite σ, σ satisfait (x_k) F ssi il existe une suite σ', différente de σ au k-ième terme, et telle que σ' satisfait F.

Elle est dite relative si la vérité ou la satisfaction sont spécifiées par rapport à un domaine d'objets, *M*, de manière à produire des clauses comme

> Pour toute suite σ dans M, σ satisfait (x_k) F dans *M* ssi il existe une suite σ', différente de σ au k-ième terme, et telle que σ' satisfait F dans *M*.

Le concept de vérité dans un modèle ou dans une structure est dû à Tarski lui-même, et la Convention T sous sa forme absolue est considérée par lui comme un cas particulier de la vérité relative. Les théories sémantiques recourant à une notion de « monde possible » (noté « *m* ») en sont des extensions. La Convention T peut alors prendre la forme :

> (T') ($\exists$ *m*) (*s* est vraie dans *m* (dans L) ssi *p* dans *m*)

Des théories conformes à la Convention (T') sont extensionnelles, puisque les conditions de vérité y sont évaluées sur des ensembles de mondes possibles. Elles respectent le principe de compositionnalité. Mais elles ne sont pas conformes à la Convention T au sens où Davidson l'entend :

> La convention T définit un but qui n'a aucun intérêt pour une grande partie des travaux contemporains en sémantique. Des théories qui caractérisent ou définissent un concept relativisé de vérité (vérité dans un modèle, vérité dans une interprétation, évaluation, ou monde possible) partent d'emblée dans une direction différente de celle qui est proposée par la Convention T. Parce qu'elles substituent un concept relationnel au prédicat de vérité à une place des phrases-T, de telles théories ne peuvent pas mener à bien la dernière étape de la récursion sur la vérité ou la satisfaction qui est essentielle au caractère de décitation des phrases-T (1973 : 68-69, *112*).

Dans toute phrase-T, la phrase citée à gauche du biconditionnel est vraie si et seulement si la phrase utilisée à droite l'est. Le prédicat de vérité nous permet ainsi de « déciter » la phrase de gauche, en l'assertant. Ainsi à partir de

> « La neige est blanche » est vraie ssi la neige est blanche

on peut déciter « la neige est blanche » pour seulement asserter

> La neige est blanche.

Si le prédicat « est vrai » devient relatif, cette opération de décitation n'est plus possible, puisque le métalangage inclut un concept, celui de modèle, d'interprétation, ou de monde possible qui n'appartient pas au langage-objet. Une objection évidente à la stratégie de Davidson sur ce point concerne le cas où le langage-objet contient des expressions intensionnelles, comme « Il est possible que » ou « il est nécessaire que ». On traduit habituellement ces expressions en quantifiant sur des mondes possibles. Comment, dans ces conditions, est-il possible de maintenir les exigences de la Convention T ?

Essayons de préciser ici l'argument de Davidson, en le comparant brièvement aux principes d'une sémantique fondée sur la théorie des modèles comme celle de Montague[1]. Dans « The Proper Treatment of

1. Montague, 1974. Pour des présentations du système de Montague, cf. Nef, 1986, Pariente et Charolles, 1990.

Quantification in English », Montague propose une théorie générale de la syntaxe et une sémantique en termes de théorie des modèles. Il traite un fragment de l'anglais contenant des quantificateurs standard et certains verbes intensionnels. Sa théorie comprend trois phases distinctes. En premier lieu, les expressions anglaises reçoivent une analyse syntaxique en termes d'une grammaire catégorielle à la Ajdukiewicz[1]. En second lieu, cette syntaxe est traduite dans la syntaxe d'une logique temporelle intensionnelle comprenant diverses constantes non logiques. Enfin, les expressions de cette logique intensionnelle reçoivent une interprétation sémantique en termes de modèles et de mondes possibles. L'interprétation procède en liant les entités linguistiques à des entités non linguistiques selon deux méthodes inspirées de Carnap, celle de l'extension et celle de l'intension. L'extension d'une expression pour un langage L est déterminée relativement à une interprétation I de L dans un monde w et à un temps t en I (*i.e* relative au modèle $< I < w, t >>$ de L). L'intension de cette expression est la signification, sens ou concept associé à l'expression. Montague définit l'intension d'une expression comme une fonction qui, pour tout monde possible w et temps t, sélectionne exactement les objets dans I qui composent l'extension de cette expression dans I à w et en t. La théorie fournit des théorèmes (simplifiés) du type suivant :

> (*Ext*) « Louisette chantonne » est vrai dans I en w et t ssi l'extension déterminée par l'intension de « Louisette » en I en w en t est un membre de l'extension déterminée par l'intension de « chantonne » en w en t.
> (*Int*) L'intension de « Louisette chantonne » dans I est une fonction complexe de l'ensemble des mondes possibles et des temps dans l'ensemble des valeurs de vérité {V, F}, fonction composée de l'intension de « Louisette » avec l'intension de « chantonne ». L'intension de « Louisette » est une fonction des mondes possibles et des temps (dans I) aux fonctions des individus dans I aux valeurs de vérité.

Montague, comme Lewis, soutient que sa sémantique a l'avantage définitif par rapport à des sémantiques « traductionnelles » de rattacher, plutôt que les expressions d'un langage aux expressions d'un autre langage, ces

1. Cf. par exemple Lewis, 1972.

expressions à des entités non linguistiques. Mais cela pose problème. La sémantique de Montague repose sur certaines hypothèses précises sur la nature de l'interprétation sémantique, et sur la manière dont la syntaxe et la sémantique se relient l'une à l'autre. Cette corrélation repose sur le principe fregéen de compositionnalité, et elle est réalisée en donnant à la syntaxe la forme d'une définition récursive simultanée des ensembles d'expressions bien formées de chaque catégorie syntaxique du langage, en construisant récursivement des expressions plus complexes à partir d'expressions plus simples, et en associant chaque règle de formation syntaxique avec une règle d'interprétation sémantique qui spécifie les interprétations des expressions constituantes. Sur ce point, la théorie de Montague ne diffère pas de celle envisagée par Davidson. La compositionnalité ne suffit pas. Il faut également définir la conséquence logique. En général elle est définie ainsi : sur un sous-ensemble K des modèles déterminés par une interprétation pour L, on définit une phrase S comme K-valide ssi elle est vraie dans chacun de ces modèles dans K. Si K est l'ensemble des modèles dans lesquels tous les mots logiques du langage reçoivent des extensions usuelles, alors ces modèles seront les modèles logiquement possibles pour L. En fait Montague établit cette restriction par un ensemble de postulats de signification, qui restreignent l'interprétation des expressions. Le problème est de savoir en quoi une sémantique de ce genre peut nous donner des « conditions de vérité » qui établissent une relation entre le langage et le monde que n'établit pas une sémantique traductionnelle. Supposons en effet que l'on énonce les mots « Louisette chantonne » et que tout ce que nous sachions du langage de l'énonciateur est que :

(*Ext'*) « Louisette chantonne » est vrai en français ssi tout ce que dénote « Louisette » est une des choses que « chantonne » dénote.
(*Int'*) La signification de « Louisette chantonne » en français est la proposition qui résulte de « Louisette » comme argument de la signification (fonction) « chantonne ».

Mais cela nous dit seulement que quand quelqu'un asserte que « Louisette chantonne » il a asserté que quelqu'un nommé « Louisette » a une expression « chantonne » qui est vrai de lui (elle). Aucune connexion directe entre les mots et ce qu'ils désignent n'est établie. Au lieu de fixer une

interprétation pour les noms et les prédicats, celle-ci est laissée ouverte. En d'autres termes, étant donné une théorie de la vérité relative, on ne peut pas dériver une théorie de la vérité absolue, du type de celle requise par Davidson. Ainsi de

(α) ($\forall$I)($\forall$p) (« Louisette chantonne » est vrai en I en p ssi l'extension de « Louisette » en I satisfait « chantonne » en I en p) (où « p » désigne des mondes possibles)

on ne peut dériver

(β) « Louisette chantonne » est vrai ssi Louisette chantonne.

Pour dériver (β) de (α), il nous faudrait une condition du type suivant :

(γ) ($\forall I$) ($\forall p$) ((l'extension de « Louisette » dans I en p = Louisette) & (x satisfait « chantonne » en I en p ssi x chantonne) & (« Louisette chantonne » est vrai dans I en p ssi « Louisette chantonne » est vrai))

Ces clauses nous disent en fait que l'on doit, pour dériver (β) de (α) savoir que l'extension de « Louisette » est Louisette, et que ($\forall x$) (x satisfait « chantonne » ssi x chantonne, et que l'interprétation I est l'interprétation dans le monde réel, et non pas dans un monde possible quelconque. Or ce sont précisément les clauses que nous donne une théorie absolue de la vérité. L'idée fondamentale d'une théorie donnant des conditions de vérité absolues est que des clauses comme

($\forall$x) (x satisfait « chantonne » ssi x chantonne)

nous fournissent une *connaissance* du sens de « chantonne » que les théories relatives ne nous fournissent pas, parce que la clause en question énonce une relation tout à fait informative sur la relation entre un mot et une propriété du monde. En ce sens, les phrases-T d'une théorie de la vérité ne sont pas triviales[1] (cf. § I.4.1).

1. Je me suis appuyé ici sur Le Pore, 1983. J.-C. Pariente m'a fait remarquer à juste titre que (γ) ci-dessus joue un rôle comparable aux postulats de signification par lesquels Montague procure aux énoncés d'ordre supérieur qu'engendre la traduction sur la logique intensionnelle des équivalents de premier ordre, et que Montague a cherché des moyens de redescendre au premier ordre. Mais le point avancé par Davidson, quand il critique la logique intensionnelle de Montague, n'est pas qu'une telle traduction serait selon lui impossible, mais le fait que la théorie de Montague sous sa forme intensionnelle ne sert pas les tâches de l'interprétation.

Décidabilité. — Comme on l'a vu, une théorie-T ne peut, pour servir de théorie de la signification, être directement appliquée à une langue naturelle. Ici encore la démarche est indirecte :

> Nous savons comment donner une théorie de la vérité pour le langage formel ; par conséquent si nous savions aussi comment transformer les phrases d'un langage naturel systématiquement en des phrases du langage formel, nous aurions une théorie de la vérité pour le langage naturel. De ce point de vue, les langages formels standard sont des techniques intermédiaires qui nous assistent pour traiter les langues naturelles comme des langues formelles plus complexes (1977 : 203, *296*).

Cette procédure peut être appelée « sémantique par procuration » (Sainsbury, 1977). A chaque phrase d'une langue naturelle, on associe une phrase d'un idiome canonique, comptant comme sa représentation sémantique. Cette représentation sémantique donne la forme logique de la phrase, c'est-à-dire révèle les traits structuraux qui affectent les conditions de vérité. A plusieurs reprises, Davidson suggère que la structure ou la forme logique révélée par une théorie de la vérité a des relations étroites avec ce que Chomsky et ses disciples appellent la « structure profonde », par opposition avec la structure de « surface », des phrases d'une langue naturelle. Davidson dit aussi qu'il veut rendre la théorie-T « décidable » ou « effective » (cf. ci-dessus la condition de scrutabilité, et 1970 : 56, *95*). Que peut-il vouloir dire par là[1] ? S'il veut dire qu'une théorie de la signification doit fournir des méthodes effectives qui modélisent ou représentent la détermination effective des significations par des locuteurs, alors cette hypothèse est extrêmement ambitieuse, si « effectif » veut dire « décidable » par une procédure effective. En effet si, d'un côté, elle signifie que les locuteurs eux-mêmes peuvent « effectivement » déterminer les significations de manière effective, la thèse de Davidson est une thèse sur la compétence sémantique qui va bien au-delà des thèses chomskyennes, car une grammaire transformationnelle classique ne nous donne qu'une énumération des associations de descriptions structurales « profondes » aux phrases « de surface » du langage concerné, et aucune mention n'est faite de méthodes effectives pour déterminer les structures profondes. Si, d'un

1. Cf. Tennant, 1977.

autre côté, la thèse de Davidson signifie qu'une théorie de la vérité tarskienne est elle-même effective, qu'est-ce que cela signifie exactement ? Veut-il dire que la récursivité d'une théorie sémantique est la même chose que sa décidabilité ? Mais c'est douteux. Un ensemble est décidable s'il y a une méthode effective pour déterminer l'appartenance d'un élément à cet ensemble. Mais nous savons, par le théorème de Church, qu'il n'y a pas de méthode effective de ce genre pour l'ensemble des formules valides de la logique du premier ordre. Or il est clair que dire qu'une théorie de la vérité est récursive ne signifie pas qu'il y a une méthode effective qui nous permette de décider, étant donné une phrase quelconque du langage dans lequel est formulée la théorie, si cette phrase est un théorème de la théorie de la vérité. La récursivité est autre chose que la décidabilité. La récursivité tient à la forme de la définition d'un prédicat important de la théorie, en l'occurrence celui de satisfaction. Dire qu'une théorie de la vérité est récursive, ce n'est pas dire plus que ceci : le prédicat de satisfaction est défini récursivement. Mais il y a une hypothèse plus plausible sur ce que veut dire Davidson. Qu'une théorie de la vérité détermine « effectivement » la signification d'une expression arbitraire veut dire qu'il y a une méthode effective pour aller d'une description structurale à une démonstration du biconditionnel tarskien approprié.

De prime abord, l'existence d'une telle méthode est compatible avec l'indécidabilité de la théorie de la vérité dans son ensemble, car l'ensemble des instances-T est seulement un fragment de la théorie. Pour voir ce que cela signifie, considérons un langage propositionnel L_3 très simple, contenant seulement trois phrases atomiques $s1$, $s2$, et $s3$, avec les connecteurs « & », « v », et «¬ » ayant leur sens usuel, et postulons :

> $s1$ signifie (dans L_3) que la neige est blanche
> $s2$ signifie (dans L_3) que la Terre tourne
> $s3$ signifie (dans L_3) que l'herbe est rouge.

Pour donner une théorie de la vérité pour L_3, on commence par faire la liste des phrases-T correspondant aux phrases de L_3 :

> (T4a) $s1$ est vrai (dans L_3) ssi la neige est blanche
> (T4b) $s2$ est vrai (dans L_3) ssi la Terre tourne
> (T4c) $s3$ est vrai (dans L_3) ssi l'herbe est rouge.

On donne ensuite pour chaque connecteur vérifonctionnel un axiome qui établit la manière dont il contribue aux conditions de vérité matérielles des phrases dans lesquelles il figure. Pour établir ceci de manière générale, on doit quantifier dans le métalangage sur les phrases du langage-objet : « σ » et « τ » désigneront ainsi respectivement les deux membres d'une phrase conjonctive ou disjonctive (où dans le métalangage les signes '&', 'v' et '¬' désignent respectivement les signes du langage-objet). On aura ainsi les trois axiomes :

(T4A) $(\forall\sigma)\,(\forall\tau)$ [‘(’ $\frown$ σ $\frown$ ‘&’ $\frown$ τ ‘)’ est vrai (dans L_3) ssi (σ est vrai (dans L_3) & τ est vrai (dans L_3))]

(T4B) $(\forall\sigma)\,(\forall\tau)$ [‘(’ $\frown$ σ $\frown$ ‘v’ $\frown$ τ $\frown$ ‘)’ est vrai (dans L_3) ssi (σ est vrai (dans L_3) v τ est vrai (dans L_3))]

(T4C) $(\forall\sigma)$ [‘(’ $\frown$ ‘¬’ $\frown$ σ $\frown$ ‘)’ est vrai (dans L_3) ssi ¬(σ est vrai (dans L_3))]

Appelons cette théorie « T_4 ». Les six axiomes propres sont adéquats pour dériver pour chaque phrase de L_3 une spécification de condition de vérité dans laquelle le côté droit des phrases-T est sans le prédicat sémantique « est vrai de ». Soit par exemple la phrase

‘(’ $\frown$ s_1 $\frown$ ‘&’ $\frown$ s_2 $\frown$ ‘)’

Par instantiation universelle à partir de (T4A) nous avons :

‘(’ $\frown$ s_1 $\frown$ ‘&’ $\frown$ s_2 $\frown$ ‘)’ est vrai (dans L_3) ssi (s_1 est vrai (dans L_3) & s_2 est vrai (dans L_3))

La logique utilisée dans le métalangage (qui est la logique ordinaire du premier ordre) nous fournit (sans l'aide des axiomes propres à T_4) toutes les instances du schéma de substitution suivant :

(A ssi B) → (Σ (A) ssi Σ(B))

où « Σ(A) » est une phrase contenant « A », et « Σ (B) » une phrase résultant du remplacement d'au moins une occurrence de « A » dans « Σ(A) » par « B ». De manière équivalente, en présence de la règle du *modus ponens*

et de la démonstration conditionnelle, la logique du métalangage donne une règle d'inférence dérivée

$$\frac{\text{A ssi B}}{\Sigma\text{(A) ssi }\Sigma\text{(B).}} \qquad \text{(inférence par substitution)}$$

En particulier, par (T4A) et le schéma de substitution, nous avons

‘(’ $\frown$ s_1 $\frown$ ‘&’ $\frown$ s_2 $\frown$ ‘)’ est vrai (dans L_3) ssi (la neige est blanche & la Terre tourne)

Il est intéressant de remarquer quelles sont les ressources déductives de la logique du métalangage. La seule règle cruciale est celle de l'inférence par substitution. Pour montrer que toutes les instances de ce schéma sont démontrables, on doit procéder par induction sur la complexité de Σ et on ferait appel aux règles d'introduction et d'élimination pour les connecteurs. Mais on n'aurait pas besoin de faire appel à la règle de double négation. En ce sens si la logique du métalangage était une logique intuitionniste, rien ne nous empêcherait de dériver les biconditionnels appropriés. La logique du métalangage repose donc sur des contraintes très faibles (cf. également ci-dessous § 5.3).

Il faut donc comprendre l'affirmation de Davidson sur le caractère « effectif » d'une théorie-T en un sens faible, distinct de celui de la décidabilité ou du caractère effectif au sens usuel. Ce qu'il veut dire essentiellement est que chaque phrase-T sera démontrable dans la théorie par une procédure spécifiée, et que la structure des démonstrations sera ce qui nous permet de « déterminer effectivement » la signification des phrases du langage. Cela a en particulier comme conséquence qu'une phrase-T *isolée* ne suffit pas à donner les conditions de vérité d'une phrase de L :

> C'est une erreur de penser que tout ce que nous pouvons apprendre d'une théorie de la vérité à propos de la signification d'une phrase individuelle est contenu dans le biconditionnel requis par la Convention T. Ce que nous pouvons apprendre est plutôt dans la démonstration d'un tel biconditionnel, car la démonstration doit établir, étape par étape, comment la valeur de vérité de chaque phrase dépend d'une structure récursivement donnée (1973 : 61, *101*).

1.4. *Forme logique, structure sémantique et extensionnalité*

Ni la Convention T ni le réquisit d'immanence ne prescrivent que le langage-objet soit « embrigadé » dans une forme logique extensionnelle du premier ordre. Néanmoins, Davidson soutient régulièrement que ces réquisits sont mieux satisfaits si le langage-objet peut recevoir une telle forme logique extensionnelle. Ceci suppose que l'on puisse systématiquement traduire le plus grand nombre de fragments de langue naturelle dans la notation canonique de la théorie de la quantification. A l'époque où Davidson a donné les premières formulations de son programme, cette hypothèse était courante chez les linguistes de l'école de la « sémantique générative » qui cherchaient à analyser la structure « profonde » des phrases en termes de leur « forme logique », assimilant celle-ci à une structure sémantique assignée directement à ces phrases[1]. Davidson n'a pas cherché à appliquer systématiquement ce programme[2], mais la démarche consistant à révéler la forme logique quantificationnelle des phrases d'une langue naturelle est néanmoins essentielle à son projet, d'une part parce que, comme on l'a vu, il est indispensable à une TS qu'elle puisse déterminer les liens inférentiels logiques entre les phrases d'un L, et d'autre part et surtout parce que le minimalisme de la procédure doit exclure le plus possible l'intensionnalité des expressions de L. Cette partie du programme de Davidson est l'une de celles qui a suscité le plus grand nombre de discussions, mais je ne la traiterai ici que brièvement[3].

Il faut d'abord préciser la notion de forme logique. Une théorie de la forme logique des phrase d'une langue naturelle est, comme on l'a vu, nécessaire en vertu même du principe de compositionnalité et du principe de vériconditionnalité : c'est parce que la sémantique des phrases est

1. Cf. notamment les articles du recueil de Davidson et Harman, 1972. J. Higginbotham, 1986, a également rattaché systématiquement le programme de Davidson en ce sens aux théories linguistiques des linguistes chomskyens.

2. Certains disciples de Davidson ont cependant cherché à étendre ses analyses à d'autres fragments, comme les adjectifs attributifs, les termes comparatifs, les termes de masse, les anaphoriques et la référence pronominale.

3. J'ai analysé certaines de ces discussions ailleurs. Cf. Engel, 1982, 1989, 1991a, 1991b. Un traitement détaillé du programme davidsonien sur ce point est fourni par Lycan, 1984, et par Davies, 1981.

donnée en termes de leurs conditions de vérité, et parce qu'elles sont construites à partir d'expressions qui affectent ces conditions de vérité, que l'on doit déterminer leur structure sémantique. Mais pourquoi cette structure devrait-elle être *logique* et non pas simplement grammaticale ? Parce que, nous dit Davidson, elle est *inférentielle* : connaître la « géographie logique » d'un langage ou d'un fragment de langage, c'est savoir quelles phrases on peut inférer d'autres phrases, et quelles relations inférentielles systématiques elles entretiennent (1969 : 94-95, *145-146*). La logique étant la théorie de l'inférence, on peut penser qu'elle nous permettra de rendre compte de ces relations. Mais, à nouveau, pourquoi penser que ces relations inférentielles seront des relations sanctionnées par la logique ? Une inférence logique est une inférence valide « en vertu de sa structure », et en vertu des expressions qui sont, dans les phrases, des « constantes logiques ». Mais comment délimiter la classe de ces expressions[1] ? Qu'est-ce qui fait, par exemple, que « Si Lassie est un chien, alors Lassie est un animal » n'est pas une inférence logique, alors que « Toto est un héros et Toto est un lâche, donc Toto est un lâche » en est une ? Une théorie de la forme logique est relative à la logique, c'est-à-dire au métalangage d'arrière-plan par rapport auquel on caractérisera les inférences comme valides « en vertu de leur forme ». Evans (1976) a bien expliqué quelle était l'originalité de la position de Davidson sur ce point. Davidson soutient que « les constantes logiques peuvent être identifiées comme les traits itératifs du langage qui requièrent une clause récursive... dans la définition de la vérité et de la satisfaction » (1973 : 71, *115*). Une inférence « structuralement valide » $S_1 \ldots S_{n-1} \vdash S_n$ est une inférence telle que le conditionnel « validant » *Si* S_1 *est vrai,...* et S_{n-1} *est vrai, alors* S_n *est vrai* est une conséquence sémantique des clauses récursives d'une théorie de la vérité (Evans, 1976 : 202). Cela ne nous donne pas une définition générale ou « transcendante » de la notion d'inférence structuralement valide, mais seulement une définition relative ou « immanente » à un métalangage particulier et à sa logique d'arrière-plan (Davidson, *ibid.*) Quelles raisons avons-nous alors de penser que la logique d'arrière-plan doit être la logique du premier ordre extensionnelle ? Pourquoi, en par-

1. Cf. Engel, 1989, § XI.3.

ticulier, cette logique ne devrait-elle pas être une logique intensionnelle, ou une logique non classique comme la logique intuitionniste ou la logique de la pertinence ? En principe, s'il l'on s'en tient au réquisit selon lequel une théorie de la vérité doit satisfaire à la Convention T, il n'y a aucune raison d'adopter un choix plutôt qu'un autre :

> La Convention T... ne fait pas mention d'extensionnalité, de vérifonctionnalité, ou de logique du premier ordre. Elle nous invite à utiliser tous les procédés possibles que nous puissions inventer de façon appropriée pour combler le fossé entre mention et usage de la phrase » (1973a : 68, *III*).

En soi, par conséquent, la seule condition est que les phrases-T soient vraies, et énoncent les conditions de vérité des phrases du langage-objet, de manière à révéler leur structure sémantique. Mais, comme on le sait, la condition d'immanence nous prescrit de nous en tenir aux théories-T homophoniques. Pourquoi devrions-nous également privilégier une représentation de la structure sémantique dans un langage extensionnel du premier-ordre ? Pas seulement pour des raisons pragmatiques de simplicité et d'économie, mais aussi parce que Davidson pense que seules les théories-T homophoniques satisfont à l'ensemble des critères formels qu'il met en avant. Mais il n'est pas évident que ce soit nécessaire, et il n'est pas évident que ce soit possible.

Pour voir en quoi ce n'est peut-être pas nécessaire, considérons brièvement la célèbre analyse que donne Davidson des phrases d'action et des modifications adverbiales (1967a)[1]. Selon cette analyse, ces phrases comportent une quantification implicite sur des événements, et les verbes d'action, ainsi que les adverbes, peuvent être traités comme des prédicats prenant une place d'argument spécifique pour des événements. Ainsi une phrase telle que « Brutus frappa César au Sénat » doit être représentée par :

(8) (∃ *e*) (Frappa (Brutus, César, *e*) & au Sénat (*e*))

On peut ainsi justifier notamment des inférences usuelles par détachement de l'adverbe, comme celle de la phrase précédente à « Brutus frappa César », c'est-à-dire de (8) à :

1. Je l'ai analysée dans Engel, 1982, 1986a, 1991a. Cf. aussi la préface de la traduction française de Davidson, 1980. Je reprends ici les principaux points de 1991a.

(9) (∃ *e*) (Frappa (Brutus, César, *e*)

L'analyse satisfait aux réquisits de Davidson : elle traite le langage-objet comme extensionnel, elle donne les conditions de vérité des phrases d'action en termes de leur structure. Elle a en outre un avantage sur lequel Davidson insiste régulièrement (cf. 1967a : 137, *197*; 1977a, et *infra*, § 6.2) : révéler l'ontologie implicite d'événements particuliers contenue dans ces expressions du langage naturel, qui se trouve par ailleurs confirmée par son analyse de l'action (Davidson, 1980, cf. *infra*, § 2.5). Davidson soutient que ces réquisits ne sont pas satisfaits par des analyses qui construisent autrement les verbes d'action (notamment comme des prédicats « polyadiques » au sens de Kenny, 1963), qui traitent les adverbes comme des modificateurs intensionnels de prédicats comparables aux opérateurs modaux comme « nécessairement » (Montague, 1974, Cresswell, 1986, Parsons, 1972), ou qui énoncent les conditions de vérité des phrases en question en termes de mondes possibles ou qui postulent une ontologie de « faits », de « situations » ou d'« états de choses » (Barwise et Perry, 1983 ; Taylor, 1985). Néanmoins l'analyse de Davidson se heurte, sur tous ces plans, à un certain nombre de difficultés bien connues : elle ne rend pas compte de certains adverbes « intensionnels » tels que « délibérément » ou « intentionnellement », elle ne fournit pas de critères précis d'individuation des événements (Davies, 1991), et néglige le fait que nombre de phrases d'action des langues naturelles semblent bien impliquer une référence à des entités telles que des « faits », ou permettre de traiter les événements non pas comme des entités particulières, mais comme des exemplifications de propriétés (Bennett, 1988). Je ne le montrerai pas ici, mais si l'on compare le cadre davidsonien et ces diverses théories (Davies, 1991), on constate qu'ils diffèrent quant à la forme logique postulée, quant à l'ontologie et quant à la nature du métalangage utilisé, mais que tous ces cadres, y compris celui de Davidson, postulent que les phrases d'action et les phrases contenant des adverbes impliquent une référence à des entités *de type événementiel* distinctes des objets, des propriétés et des relations, et qu'ils permettent tous de rendre compte de la structure sémantique compositionnelle et véri conditionnelle de ces phrases. De plus tous ces cadres rencontrent des difficultés qui leur sont propres, relativement à

l'ontologie postulée dans chaque cas et aux conditions d'individuation des entités postulées. Si c'est le cas, qu'est-ce qui nous permet de choisir une ontologie plutôt qu'une autre et un métalangage plutôt qu'un autre ? On peut dire que si la condition minimale que doit remplir une théorie sémantique systématique est avant tout celle de spécifier la structure sémantique en termes des conditions de vérité, alors cette condition est remplie quels que soient l'ontologie, le mode de représentation logique de la structure en question, et la nature du métalangage employé. Comme le dit Davies, « les réquisits d'adéquation descriptive d'une théorie sémantique systématique imposent peu de contraintes sur la métaphysique des événements sous-jacente à la théorie » (1991 : 79). Il s'ensuit que le format particulier préconisé par Davidson pour une théorie sémantique — le format d'une théorie-T extensionnelle représentant ces phrases dans une forme logique extensionnelle — n'a par lui-même, quand il s'agit de décrire adéquatement la sémantique d'un fragment de langue naturelle comme celui des phrases d'action, rien de contraignant. On doit donc dire que, sur ce plan, l'analyse proposée par Davidson est possible, mais qu'elle ne s'impose pas comme la seule possible.

Ce fait, à lui seul, limite sérieusement la prétention du programme de Davidson à s'imposer comme un cadre nécessaire d'analyse sémantique des langues naturelles. Mais cela n'infirme pas pour autant les prétentions de ce programme à constituer un cadre descriptif suffisant pour une sémantique. Ce qui, en revanche, peut infirmer ces prétentions serait le fait qu'une certaine catégorie d'expressions ne peuvent pas être analysées au moyen d'une théorie-T extensionnelle :

> Bien entendu mon projet requiert bien qu'on puisse soumettre toutes les phrases des langues naturelles à une théorie-T, et donc, si les idiomes intensionnels résistent à un tel traitement, mon projet s'effondre (1976 : 176, 259).

Sur ce point, la difficulté principale à laquelle se heurte Davidson est qu'un grande nombre d'expressions des langues naturelles semblent être intensionnelles. Ce n'est pas le cas seulement des adverbes tels que « nécessairement » ou « intentionnellement », mais également, et notoirement, de toutes les phrases rapportant des contenus d'attitudes propositionnelles, de la forme « X croit que... », « X désire que... » ou « X dit que... ».

Ces phrases violent les réquisits usuels de l'extensionnalité (substitution *salva veritate* des termes coréférentiels, quantification à l'intérieur des contextes gouvernés par les verbes d'attitudes). Comment, dans ce cas, les conditions d'une analyse extensionnelle de leur forme logique peuvent-elles être satisfaites ?

Davidson s'attaque à ce problème dans « On Saying That » (1968). Il stipule que toute théorie sémantique adéquate des phrases rapportant des contenus d'attitudes propositionnelles doit donner la forme logique de ces expressions — et par conséquent expliquer pourquoi les inférences usuelles par substitution de termes coréférentiels ou par quantification dans ces contextes sont invalides —, éviter de recourir à des entités intensionnelles telles que des « propositions », et rendre compte d'une intuition tenace : que le sens d'une phrase « *p* » dans un contexte tel que « X croit que *p* » doit être le même que celui qu'elle a dans un contexte extensionnel ordinaire, à savoir quand elle n'est pas enchâssée dans la portée d'un verbe tel que « croit que » ou « dit que ». Les théories usuelles, qui, comme celle de Frege, postulent que la forme logique de ces phrases implique une relation à certaines entités telles que les sens des phrases, ou des propositions, violent ce réquisit d'« innocence sémantique », puisqu'elles disent que le sens de ces phrases différera de celui qu'elles ont dans les contextes ordinaires, en même temps qu'elles violent le second réquisit, puisque les sens ou les propositions sont des entités intensionnelles. Comme on l'a vu (§ 1.1), ce n'est pas tant parce qu'elles postulent ces entités que ces théories sont inadéquates, que parce qu'elles ne montrent pas clairement comment le sens des expressions composant une phrase de ce type dépend du sens de ses parties. Davidson propose que des phrases au discours indirect telles que

(10) Galilée dit que la Terre tourne

soient paraphrasées ainsi :

(11) Galilée dit ceci. La Terre tourne

Dans cette paraphrase, la conjonction de subordination « que » est traitée comme un démonstratif (« ceci »), qui désigne une énonciation de la phrase « La Terre tourne » par un locuteur (Davidson s'aide du fait qu'en anglais,

that peut être à la fois un démonstratif et une conjonction de subordination)[1]. La première phrase (« Galilée dit ceci ») représente l'énonciation de Galilée et celle du locuteur qui la rapporte comme étant en relation de « mêmedire » avec ce que désigne le démonstratif. En d'autres termes, en rapportant ce qu'un locuteur dit, nous ne faisons pas une affirmation, mais deux : l'une d'après laquelle Galilée a dit quelque chose, l'autre qui rapporte (dans notre propre langage) ce que Galilée a dit et par laquelle nous énonçons que ce que Galilée a dit et ce que nous avons dit est « la même chose ». C'est une parataxe, d'où l'appellation de la théorie comme « théorie parataxique du discours indirect ». Notons que le recours à l'idée de « mêmedire » ne revient nullement à réintroduire des contenus de propositions qu'on cherchait à éviter : Davidson entend seulement stipuler une exigence *conceptuelle* que doit satisfaire la phrase énoncée et la phrase rapportée. Il ne dit pas que la relation de « même-dire » figure dans la forme logique de la phrase (II). L'analyse parataxique peut s'appliquer à des phrases gouvernées par des verbes comme « croire » , « désirer » et d'autres verbes d'attitudes. Elle implique que la forme logique des phrases au discours indirect ou rapportant des contenus d'attitudes propositionnelles soit relationnelle, mais que la relation en jeu ne soit pas une relation à des propositions ou à des contenus (ou sens, ou significations) de phrases, mais à des *énonciations* (réelles ou potentielles) des phrases correspondantes. L'intensionnalité ou l'« opacité » des phrases enchâssées dans des verbes d'attitudes s'explique alors comme un phénomène indépendant de la prétendue non-extensionnalité de ces phrases. Les deux phrases dans (II) sont toutes deux extensionnelles, et les mots y ont la même signification que dans les usages ordinaires (en dehors des contextes où on rapporte ce qui est dit, cru, etc.). On ne peut cependant appliquer les principes de substitution et de généralisation existentielle à la seconde phrase. Cela n'a rien à voir avec la non-extensionnalité de ces phrases, mais seulement avec le fait que la référence du démonstratif « ceci » cesse d'être la même. La solution de Davidson appartient à la catégorie des analyses « citationnelles » du discours indirect, selon lesquelles rapporter le contenu d'une phrase c'est la citer. Mais contrairement aux versions courantes de ces analyses, elle ne se heurte pas à la difficulté relevée par Church (1950) selon laquelle la traduction dans un autre langage de la phrase rapportée modiferait le compte rendu.

1. Mais ce fait n'est pas essentiel à son analyse. Cf. ma note, p. 161, de la traduction française.

Cette analyse est tout à fait essentielle au programme sémantique de Davidson, non seulement pour les raisons que nous venons de mentionner, mais aussi parce que, comme on va le voir dans la section suivante, elle suppose qu'il est possible de rapporter le contenu des assertions et des croyances d'un locuteur *sans* savoir si les phrases que ce locuteur tient pour vraies *sont* vraies, et sans savoir quelles sont leurs conditions de vérité ni leur signification. Or on peut se demander si rapporter le discours ou les croyances d'autrui est possible sans connaître le sens de ce qu'il dit ou croit. Nous rencontrerons cette difficulté à de nombreuses reprises dans ce qui suit. Mais on supposera pour le moment qu'elle peut être être résolue.

D'un point de vue strictement sémantique, cette analyse ingénieuse se heurte aussi à des difficultés familières, dont les principales sont les suivantes[1]. En premier lieu, elle est contre-intuitive : notre intuition nous dit que la syntaxe des phrases rapportant des attitudes n'est pas composite. En second lieu, cette analyse suppose que la phrase rapportée soit énoncée, au moins potentiellement. Mais nous pouvons rapporter des contenus d'attitudes qui ne font jamais, même potentiellement, l'objet d'une énonciation de la part d'un locuteur. En troisième lieu, l'analyse ne semble pas pouvoir s'appliquer à des phrases au discours indirect itérées, telles que « Galilée a dit que Copernic avait dit que... » parce que, dans l'hypothèse où « que » fonctionnerait ici comme un démonstratif le premier « que » ne peut pas faire référence à ce qui serait contenu dans les guillemets « ... ». Sans entrer ici dans ces difficultés et sans présupposer non plus qu'elles ne puissent pas être résolues, on doit admettre qu'elles limitent à nouveau sérieusement la thèse de Davidson selon laquelle il est possible de soumettre les « idiomes intensionnels » à un traitement extensionnel.

Ce que ces difficultés, comme celles qui se posent pour l'analyse sémantique d'autres expressions, indiquent, est que le programme consistant

1. Cf. en particulier Baldwin, 1982, Burge, 1986, Schiffer, 1987 : 122-138. Toutes ces objections sont bien résumées par Seymour, 1994.

à imposer un format unique, celui d'une théorie-T, à l'ensemble d'une théorie sémantique pour une langue donnée n'est pas assuré de réussir. Rien ne nous garantit qu'il existe un type d'interprétation sémantique uniforme pour toutes les expressions d'une langue naturelle, et qu'on ne doive pas spécifier, pour chaque catégorie d'expressions discernées par une syntaxe adéquate, des règles d'interprétation particulières à ces types d'expressions, qu'il s'agisse, par exemple, des adverbes, des noms, des termes de masse, ou des adjectifs attributifs, selon le principe de ce que Evans (1976) appelle une « sémantique interprétationnelle ». Une sémantique non uniforme de ce genre ne violera cependant pas les réquisits minimaux que Davidson impose à une théorie de la structure sémantique : elle restera compositionnelle, vériconditionnelle, et permettra d'assigner aux expressions des formes logiques qui rendront compte des inférences particulières validées par ces types d'expression. Mais elle perdra aussi l'homogénéité de la proposition de Davidson : elle pourra user de ressources intensionnelles, et par conséquent modifier le format initialement proposé.

Nous ne devons cependant pas conclure que ces problèmes menacent directement le projet davidsonien dans son ensemble, parce que, comme on va le voir, il ne revient pas à soutenir qu'une théorie-T extensionnelle serait *par elle-même* une théorie de la signification, ni qu'une telle théorie doive se passer intégralement de notions intensionnelles.

1.5. Une théorie de la vérité est-elle une théorie de la signification ?

Venons-en à présent au problème fondamental posé par la proposition de Davidson : sa thèse majeure est qu'une théorie de la signification pour une langue naturelle doit prendre la forme d'une théorie-T satisfaisant aux conditions formelles que nous venons d'énumérer. Cela peut-il vouloir dire qu'une théorie de la *signification est*, ou *s'identifie à*, une théorie-T ? Dans certains textes, Davidson semble le suggérer, en identifiant en particulier (i) donner la signification d'une phrase et donner ses conditions de vérité et (ii) donner les conditions de vérité d'une phrase et déterminer sa structure sémantique, et il semble en inférer que puisqu'une

théorie-T donne, par l'intermédiaire de ses phrases-T, les conditions de vérité des phrases de *L*, et par l'intermédiaire de sa structure récursive leur structure sémantique, une théorie-T n'est rien d'autre qu'une théorie de la signification et que connaître une théorie-T pour *L* c'est être capable de comprendre L :

> Il n'est bien entendu pas nécessaire de supprimer le lien évident qui existe entre une définition de la vérité du type de celle que Tarski nous a montré comment construire, et le concept de signification. Ce lien est le suivant : la définition se construit en donnant des conditions nécessaires et suffisantes pour la vérité de chaque phrase, et donner des conditions de vérité est une manière de donner la signification d'une phrase. Connaître le concept sémantique de vérité pour un langage, c'est savoir ce que c'est pour une phrase — n'importe quelle phrase — que d'être vraie, et ceci revient, *en un sens important que nous pouvons donner à l'expression, à comprendre le langage* [mes italiques] (1967 : 24, *51*).

Une bonne partie, sinon l'essentiel, du problème tourne autour de la question de savoir ce que veulent dire ici des expressions comme « donner les conditions de vérité » et « connaître les conditions de vérité » d'une phrase. Tant qu'on n'a pas répondu à ces questions on ne peut pas espérer avoir clarifié, et *a fortiori* justifié, le lien entre signification et vérité. Mais Davidson dit qu'avec une théorie-T on a une réponse. Laquelle ?

La réponse « officielle » de Davidson est, comme l'a vu, la suivante. Connaître la signification d'une phrase *s* d'une langue naturelle *L*, c'est connaître ses conditions de vérité telles qu'elles sont exprimées par une phrase-T, si deux conditions sont remplies : (a) cette phrase-T est vraie (b) cette phrase-T est une conséquence (un théorème) d'une théorie-T axiomatisée de manière finie qui implique conjointement une phrase-T vraie pour toute phrase d'un langage, et permettant de discerner la structure de ces phrases. Mais il nous faut encore comprendre comment ces conditions sont supposées fonctionner. Ignorer leur nature exacte peut conduire à deux sortes d'objections usuelles : qu'une théorie-T est triviale, et par conséquent ne nous apprend rien sur la signification, et qu'elle est, sinon triviale, inadéquate.

Commençons par l'objection de trivialité[1]. Une théorie-T établit les significations des phrases d'un langage L en donnant leurs conditions de

1. Formulée notamment par Stich, 1975, Lycan, 1984, chap. 2.

vérité, au moyen des phrases-T qui en sont les conséquences. Dans une phrase-T, la phrase située à droite du biconditionnel énonce les conditions de vérité de la phrase située à gauche. Dans l'exemple canonique

(12) « La neige est blanche » est vrai ssi la neige est blanche

la condition de vérité de « la neige est blanche » est que la neige est blanche. Mais si cela semble manifestement correct cela paraît également trivial, ce qui est la source d'une des objections les plus répandues contre l'idée qu'une phrase-T pourrait nous donner la signification de la phrase mentionnée. Comme on a déjà eu l'occasion de le souligner, les phrases-T comme (12) paraissent triviales parce que nous en comprenons le métalangage, c'est-à-dire parce que nous savons ce que signifie « la neige est blanche » en français. Mais on peut s'assurer aisément que ce n'est pas le cas, si par exemple on considère une théorie de la vérité en français pour des phrases allemandes. Dans ce cas,

(12') « Der Schnee ist weiss » est vrai ssi la neige est blanche

est une phrase-T informative. Mais nous voulons pouvoir nous en tenir au cas homophonique, et donc accepter des phrases comme (12). Si (12) paraît triviale, c'est parce que l'on suppose qu'elle établit une relation que la phrase « la neige est blanche » a avec elle-même, alors que la phrase située à droite du biconditionnel ne mentionne pas la phrase à gauche, mais l'utilise. Ce faisant, (12) établit au moins une propriété sémantique de la phrase, sa vérité. En ce sens, (12) établit un fait contingent propre à la phrase « la neige est blanche », susceptible d'être appris, et donnant une information sur une condition dans laquelle la phrase concernée est vraie. Ce qui obscurcit ce fait est qu'il nous semble que quelqu'un qui ignorerait tout du français pourrait, en prenant toutes les phrases qu'il rencontre et en les citant au moyen du prédicat de vérité, établir la liste des phrases-T, et par conséquent la théorie-T correspondante, sans pour autant connaître la signification de ces phrases. On pourrait ainsi produire des phrases-T comme :

(13) « Un snark est un boojum » est vrai ssi un snark est un boojum.

sans éclairer pour autant la signification des phrases concernées. Il est clair que tant que l'on reste sous l'illusion qu'on a simplement établi une relation entre une phrase et elle-même (« elle est vraie si elle est vraie »), rien ne peut sortir des phrases-T. La situation serait en ce sens exactement identique à celle de la sémantique traductionnelle, dont on peut connaître les clauses sans savoir ce qu'elles signifient. Il est exact que, de ce point de vue, savoir que la phrase-T (13) est vraie est aussi peu informatif que savoir que

(13') « Un snark est un boojum » signifie qu'un snark est un boojum.

Mais assimiler les deux repose sur une illusion, et le fait que la phrase impliquée par (13) soit un non-sens par nos critères usuels n'a pas ici de pertinence. Quelqu'un qui saurait seulement que (13) est vrai (par exemple s'il a entendu quelqu'un qui est une autorité fiable asserter la phrase correspondante) ne saurait évidemment pas pour autant ce que « un snark est un boojum » veut dire. Ce n'est pas la même chose de savoir que (13) est une phrase vraie et de savoir que « un snark est un boojum » est vraie si et seulement si un snark est un boojum. En d'autres termes, quelqu'un qui connaît sait que (13) est une phrase vraie ne sait pas nécessairement ce que (13) exprime, ou quel est son contenu, et c'est ce dernier savoir qui rend (13) informatif. Comme le souligne Dummett sur ce point[1], la distinction dont nous avons besoin est la distinction entre savoir qu'une phrase est vraie et savoir quelle proposition cette phrase exprime. Cette distinction, selon Dummett, ne nous engage en rien à reconnaître l'existence d'entités suspectes telles que des « propositions » ou des « significations », mais nous engage seulement à distinguer savoir quelque chose, et savoir qu'une certaine phrase est vraie. Davidson n'a aucune raison de la refuser. La refuser reviendrait à assimiler le savoir exprimé par les phrases-T, dont Davidson soutient qu'il est constitutif de notre connaissance d'un langage, avec la seule connaissance de la vérité des phrases-T. Comme on va le voir, cette dernière connaissance est essentielle, pour un interprète du langage, pour l'attribution à un locuteur du savoir exprimé par les phrases-T. Mais il ne s'ensuit pas que ce que

1. Dummett, 1975 : 105-106. Je reviens sur ce raisonnement au chapitre 5.

connaît un locuteur, quand il connaît son langage, soit seulement que les phrases-T sont vraies.

Mais en quoi consiste la connaissance du fait que « la neige est blanche » est vrai si et seulement si la neige est blanche ou de la proposition que cette phrase exprime, qui puisse valoir comme une connaissance de la signification de la phrase « la neige est blanche » ? Nous pouvons nous demander de quelle sorte de connaissance il s'agit, c'est-à-dire si c'est une connaissance propositionnelle, supposée être articulable par un locuteur, ou s'il s'agit d'une autre forme de connaissance, tacite ou implicite, qui relèverait d'un savoir « comment » plutôt que d'un « savoir que ». Mais je déférerai l'examen de ce point au chapitre 7. Quoi qu'il en soit, la réponse est qu'il n'y a aucun « fait », aucune pièce particulière de savoir qui compte comme une connaissance isolée des conditions de vérité de la phrase en question. Par définition, comme on l'a déjà vu à la fin de la section précédente, le locuteur n'a pas à connaître des phrases-T isolées, et une théorie de la signification n'est pas contenue dans l'énoncé de phrases-T isolées. En premier lieu, un locuteur qui sait ce que signifie « la neige est blanche » sait ce que signifient un grand nombre d'autres phrases. En second lieu, ce qu'il sait n'est pas contenu dans la seule phrase-T correspondante, mais, comme on l'a vu, dans la démonstration de cette phrase, qui doit, dans une théorie-T, révéler récursivement sa structure à partir des éléments qui la composent. « Connaître la signification » d'une phrase, et, selon l'hypothèse présente, connaître ses conditions de vérité, c'est connaître la signification et les conditions de vérité d'un grand nombre d'autres phrases, et c'est également connaître la manière dont on peut dériver ces conditions de vérité. Supposer le contraire, c'est ignorer le holisme de la signification. Les phrases-T n'apparaissent comme triviales que si on les considère isolément, et si l'on ne tient pas compte de la structure récursive de la théorie-T dont elles font partie, qui ne peut non plus, en ce sens, consister en une liste de phrases-T indépendantes les unes des autres.

La même condition holistique permet de répondre à la seconde objection, celle de l'inadéquation d'une théorie-T comme théorie de la signification. Il y a deux versions de cette objection. La première consiste à remarquer que si, selon la conception vériconditionnelle, la signification

d'une phrase s'identifie à ses conditions de vérité, alors deux phrases qui ont les mêmes conditions de vérité ou qui sont extensionnellement équivalentes devraient avoir la même signification. Par exemple « p » et « q » et « non (non p ou non q) » devraient avoir la même signification. Or quelqu'un peut bien savoir que la première est vraie sans savoir que la seconde l'est. Mais une théorie-T rend compte de cette différence. Les conditions de vérité données par les phrases-T correspondantes ne sont pas les mêmes :

(a) « p et q » est vrai ssi « p » est vrai et « q » est vrai
(b) « non (non p ou non q) » est vrai ssi il n'est pas vrai que p n'est pas vrai ou que q n'est pas vrai

et les deux phrases-T ne seraient pas démontrées de la même manière. Il est donc faux qu'une théorie-T ne puisse pas représenter ces différences de signification[1]. La seconde version de l'objection est la suivante. Si tout ce qu'exigeait Davidson est que les phrases-T soient vraies, il suffirait que la phrase située à gauche du biconditionnel matériel « ssi » ait la même valeur de vérité que la phrase située à droite. En ce sens, pourquoi une théorie qui produirait des phrases-T comme :

(14) « La neige est blanche » est vrai ssi l'herbe est verte

serait-elle inadéquate ? Nous ne pouvons pas, pour rejeter (14), faire appel ici à un savoir quelconque concernant la signification des deux phrases (en l'occurrence le fait que celle de gauche ne traduit pas celle de droite), puisque par définition c'est la phrase-T (14) qui est supposée nous fournir ce savoir. La réponse de Davidson est la suivante :

> Le caractère grotesque de (14) ne révèle rien en lui-même contre une théorie qui a cette phrase pour conséquence, pour autant que la théorie donne les résultats corrects, pour toute phrase (sur la base de sa structure, car il n'y a pas d'autre moyen). Il n'est pas facile de voir comment (14) pourrait faire partie d'une telle entreprise, mais si elle en faisait partie — si, en d'autres termes, (14) découlait d'une caractérisation du prédicat « est vrai » qui conduisait à une association

1. L'objection est celle, par exemple, de Harman, 1974, et 1986. J'ai commenté cet exemple dans Engel, 1989 : 55-57.

> inévitable de vérités avec des vérités, et de fausstés — avec des faussetés —, alors il ne resterait rien, à mon sens, qui soit essentiel à la notion de signification à représenter » (1967 : 26, 53).

Cette réponse est identique à la précédente. Notre connaissance de la signification d'une phrase n'est pas isolable de notre connaissance de la signification d'autres phrases, et par conséquent, si cette signification est caractérisée par une théorie-T, une phrase-T n'est jamais *à elle seule* la transcription de notre savoir concernant la signification. En d'autres termes, la condition (a) ci-dessus — que les phrases-T soient vraies — ne peut pas fonctionner sans la condition (b), qu'elles soient dérivées d'une théorie-T récursive. Par suite, ou bien (14) est une conséquence correcte d'une telle théorie, et dans ce cas (14) transcrit bien les conditions de vérité de « La neige est blanche », même si celles-ci sont bizarres, ou bien elle n'est pas une conséquence d'une telle théorie, et en ce cas elle est incorrecte. Ce qui peut, en l'occurrence, nous amener à penser que (14) n'est pas une conséquence d'une théorie de la vérité correcte et que si (14) était confrontée à un ensemble d'autres phrases-T nous découvririons, à un moment ou à un autre, que cette phrase ne peut pas en être la conséquence[1]. Comme le disent Evans et McDowell :

> Il est difficile de voir en quoi une théorie qui aurait (14) comme conséquence serait capable d'un accouplement infini de vérités du langage-objet avec des vérités du métalangage et des faussetés du langage objet avec des faussetés du métalangage. (14) ne pourrait résulter que d'une connexion sémantique de « blanc » avec des choses vertes et de « neige » avec de l'herbe, qui, bien qu'étonnamment hasardeuse ici, serait ici de nature à nous induire en erreur (extensionnellement parlant) quelque part ailleurs (Evans et McDowell 1976, xiv).

En d'autres termes, on peut espérer que le fait que la condition holistique qui pèse sur une théorie de la signification pourrait nous permettre d'exclure des phrases-T grotesques comme (14). Mais ce holisme est-il suffisant ? Pourquoi un locuteur ou une communauté de locuteurs ne parleraient-ils pas un langage dont une théorie-T produirait des phrases comme (14)

1. J'ai explicité ce raisonnement sur un exemple, *in* Engel, 1989 : 147-148. En fait, comme nous le verrons aux chapitres suivants, on a toutes les raisons de supposer que l'interprète d'un langage sera confronté *très souvent* à des phrases-T comme (12).

comme conséquences ? Nous trouvons (14) bizarre parce que nous ne croyons pas un instant que « la neige est blanche » est vrai si et seulement si l'herbe est verte, et *a fortiori* nous n'envisageons pas un instant non plus de créditer les locuteurs d'une connaissance de la proposition exprimée par (14). Mais qu'est-ce qui nous permet d'exclure que des locuteurs aient ce genre de croyances ? Peut-être ces locuteurs vivent-ils dans un environnement tel que c'est effectivement le cas que la neige est blanche si et seulement si l'herbe est verte, ou peut-être vivent-ils dans la croyance erronée que la neige est de l'herbe. Autrement dit, comment pouvons-nous nous assurer que nos attributions de signification sont correctes tant que nous ne faisons pas appel aux croyances des locuteurs, à nos propres croyances, et à la nature de l'environnement ou de la réalité ? La condition holistique ne suffit pas elle seule, et par conséquent les conditions (a) et (b) de Davidson ne suffisent pas non plus si l'on ne dit pas comment la vérité des phrases-T est vérifiée.

Pour s'en assurer, on peut considérer une objection, due à John Foster (1976), destinée à montrer que les conditions « formelles » (a) et (b) sont insuffisantes pour qu'une théorie-T soit une théorie de la signification. Foster nous invite à considérer une théorie-T pour un langage L_4 dont le vocabulaire serait le suivant :

(i) un prédicat à deux places « P »
(ii) le connecteur barre de Scheffer « / » avec son sens usuel
(iii) le quantificateur universel « $\forall$ » avec son sens usuel
(iv) un stock dénombrable de variables individuelles : $x_1, x_2, x_3, \ldots$
(v) les parenthèses « (» et «) »

ainsi que toutes les phrases formées à partir de ces symboles, selon les règles usuelles du calcul des prédicats standard. On stipule la clause usuelle pour l'assignation de suites aux variables :

(a) pour toute suite σ, σ^* (x_k) = le k-ième membre de σ
(où '*' est la fonction qui assigne une valeur à la variable v_k pour la suite).

Supposons que l'interprétation de « P » soit le prédicat « est une partie de ». Pour une phrase ouverte « P (x_k, y_l) », la clause correspondante est :

(b) Pour tout σ, σ satisfait $P\ \cap(\cap x_k \cap,\ \cap y_l \cap)$ ssi $\sigma^*(x_k, y_l)$ est P (*i.e.* ssi l'objet que σ assigne à la première variable est dans la relation P (est une partie de) avec l'objet que σ assigne à la seconde variable)

puis pour les clauses plus complexes :

(3) Pour tout σ satisfait $A \cap / \cap B$ ssi σ ne satisfait pas A ou B

(4) Pour tout, σ, σ satisfait $\forall \cap x_k \cap \forall \cap x_{yi} \cap P \cap(\cap x_k,\ \cap y_i \cap)$ ssi pour toute suite $\sigma' \approx \sigma$ satisfait P

la vérité étant finalement définie comme la satisfaction par toutes les suites, selon la procédure tarskienne usuelle[1].

Cette théorie-T (appelons-la Θ) donne une caractérisation adéquate de la vérité pour $L4$. Mais elle ne donne pas une caractérisation adéquate de la signification, parce que pour chaque phrase de $L4$, il y a une infinité d'interprétations compatibles avec Θ. On peut le voir en considérant la clause (2), qui nous dit que P s'applique à toutes les paires d'objets et seulement à elles qui sont reliées par une relation de partie à tout. Mais cette interprétation est purement extensionnelle. Elle s'applique à tout prédicat P' qui ait la même extension, même s'il n'a pas la même intension, ou signification, que P.

Supposons en effet que P' soit le prédicat « La Terre tourne et ...est une partie de ... » où « La Terre tourne » est une phrase vraie. En ce cas, P et P' sont coextensifs ; mais ils n'ont pas la même signification. Il y a, comme le souligne Foster, une infinité d'exemples de ce type, soit que l'on invente des prédicats coextensifs, soit que l'on invente des connecteurs coextensifs. Soit Θ' la théorie-T caractérisant le langage $L4'$ contenant P' au lieu de P. Θ et Θ' sont coextensives ; elles préservent la syntaxe de leurs langages respectifs et sont toutes deux des théories-T adéquates ; mais aucune d'elles ne caractérise adéquatement le langage dont elle est la théorie, puisqu'une multiplicité d'interprétations sont compatibles avec l'une ou l'autre d'entre elles. Nous ne pouvons pas faire appel ici au fait que les phrases-T détermineraient la signification sur la totalité des assignations de conditions de vérité pour soutenir qu'elles déterminent la signification des phrases de $L4$ ou de $L4'$ puisque par définition elles le font, mais sous-déterminent les assignations possibles de signification.

1. Pour un exposé de cette procédure, cf. Engel, 1989, chap. 5.

La conclusion qui s'impose est qu'il est faux de dire qu'une théorie de la vérité est, par elle-même, et en vertu des réquisits formels de Davidson, une théorie de la signification. On peut en conclure deux choses : ou bien, si l'on veut préserver la proposition selon laquelle une théorie de la signification doit être une théorie de la vérité, qu'il faut réviser les conditions formelles très strictes que Davidson impose à une théorie de la vérité, peut-être en cherchant comment on peut formuler des théories de la vérité qui utilisent la notion de signification, ou bien que le format imposé par Davidson peut être conservé, mais doit être assorti d'autres sortes de conditions.

J'ai suggéré ci-dessus comment on pouvait envisager la première hypothèse. Mais en tout état de cause, ce n'est pas celle que Davidson favorise. Dans sa réponse à Foster, il soutient qu'il n'a jamais dit qu'une théorie-T était, *par elle-même*, une théorie de la signification : « Une théorie de la vérité, qu'elle soit ou non bien choisie, n'est pas une théorie de la signification » (1976 : 179, *262*)[1]. Il n'a jamais dit autre chose parce qu'il soutient que théorie-T n'est une théorie de la signification que moyennant certaines conditions empiriques qui pèsent sur la manière dont nous devons interpréter un langage. Comme l'exemple de (14) ci-dessus nous le laisse entrevoir, ces conditions empiriques ont à voir avec la manière dont nous attribuons, outre des significations aux phrases du langage d'un locuteur ou d'une communauté, certaines croyances à ce locuteur ou à cette communauté, conjointement avec les croyances que nous avons nous-mêmes. Ces conditions sont celles de l'interprétation radicale.

Si une théorie-T n'est pas une théorie de la signification, on ne peut pas identifier « donner la signification » d'une phrase avec « donner ses conditions de vérité », comme semblait le soutenir Davidson dans le passage de 1967 cité plus haut (§ 1.4). Il s'ensuit que Davidson ne soutient pas la version forte (§ 1.2) du principe de vériconditionnalité, mais

1. Davidson est également très clair sur ce point dans ses Dewey Lectures : « Mon erreur [dans « Truth and Meaning »] était de penser que nous pourrions *à la fois* prendre une définition de la vérité à la Tarski comme nous disant tout ce que nous avons besoin de savoir quant à la vérité *et* utiliser la définition pour décrire un langage » (1990 : 286 : note 20).

sa version faible. En d'autres termes, il soutient que l'on peut inférer de

s signifie (dans *L*) que *p*

que :

(T) *s* est vrai (dans *L*) ssi *p* (moyennant les conditions formelles d'une théorie-T)

mais il ne soutient pas qu'on puisse inférer conversement de (T) que *s* signifie (dans *L*) que *p*. La signification d'une phrase détermine les conditions de vérité qu'elle a, mais les conditions de vérité ne déterminent pas la signification[1].

Comment peut-on alors concevoir la relation entre une théorie de la vérité et une théorie de la signification ? McDowell (1977) a fourni une analyse éclairante de ce lien en comparant la proposition de Davidson avec certaines des idées fondamentales de Frege. Frege disait que les sens des phrases peuvent être spécifiés en donnant leurs conditions de vérité, et que le sens d'une partie composante d'une phrase est sa contribution aux sens des phrases dans lesquelles elle figure. En d'autres termes, il disait qu'une théorie du sens doit à la fois respecter le principe de vériconditionnalité et celui de compositionnalité[2]. On peut adapter ces deux principes à une certaine conception d'une théorie du sens : celle-ci assigne une certaine propriété, que l'on peut appeler valeur sémantique à chaque constituant simple d'une phrase formée selon la syntaxe du langage, et elle établit des règles qui déterminent la valeur sémantique des expressions complexes, étant donné la valeur sémantique de leurs composantes. Si nous tenons la valeur sémantique propre à chaque phrase comme la propriété que cette phrase a d'être vraie dans certaines conditions, nous pouvons dire qu'une théorie du sens pour un langage nous montre comment dériver, pour toute phrase du langage, un théorème de la forme

1. Loewer et Le Pore (1989, 1989a) ont soutenu que Davidson « aurait dû » répondre à Foster en invoquant sa théorie du discours indirect. Mais il n'a pas besoin, à mon sens, de faire dépendre son analyse de cette théorie, si les remarques qui précèdent sont correctes.

2. C'est le passage de Frege, 1893 I.32 (cité en introduction ci-dessus) qui est invoqué. Pour une analyse de ce passage, cf. Engel, 1985 : 68-69. J'ignore ici le fait que le « sens » désigne chez Frege un certain type d'entité.

« *'s'* est vrai ssi *p* ». En d'autres termes, la suggestion est qu'une théorie du sens soit une théorie de la vérité. Ceci correspond à la proposition de Davidson sous la forme radicale où on peut l'interpréter à partir des déclarations de « Truth and Meaning ». Mais nous avons vu précisément, avec l'objection de Foster, qu'une théorie de la vérité n'était pas une théorie du sens. McDowell suggère qu'elle puisse néanmoins *faire office* de théorie du sens, sans s'identifier pour autant à cette dernière :

> Faire *office* de théorie du sens n'est pas la même chose qu'en être une, selon une certaine conception stricte de ce que c'est qu'en être une. Il était clair en tout cas qu'une théorie du type de celle que Davidson envisage n'établit pas, en disant ce qu'elle fait, les sens des expressions (McDowell 1976 : 161).

Comment alors une théorie de la vérité peut-elle jouer ce rôle de substitut d'une théorie du sens ? Prenons l'exemple, utilisé par McDowell, d'une théorie de la signification pour un langage contenant des noms propres. La distinction fregéenne entre le sens et la référence est destinée notamment à résoudre le fameux problème du sens des énoncés d'identité contenant des noms propres : comment des énoncés comme « Hespéros = Phosphoros » peuvent-ils différer en valeur de connaissance d'énoncés comme « Hespéros = Hespéros », alors qu'ils ont les mêmes références et valeur de vérité ? En raison, soutient Frege, du sens des noms propres « Hespéros » et « Phosphoros », qui n'apportent pas la même contribution sémantique dans chaque phrase. La conception fregéenne nous enjoint donc de doter chaque nom d'un sens spécifique. Elle s'oppose ainsi à toute théorie qui, comme celle de Stuart Mill (ou ses variantes contemporaines), soutiendrait que la signification des noms propres réside seulement dans leur dénotation. Faut-il choisir entre l'une ou l'autre de ces théories du « sens » des noms propres ? La très vaste littérature contemporaine sur cette controverse porte sur les raisons de ce choix[1]. Mais si nous adoptons la conception proposée par McDowell, nous pouvons éviter, ou en tout cas différer, ce choix. Dans une théorie de la vérité pour un langage contenant les noms « Hespéros » et « Phosphoros », nous aurions des axiomes du type suivant :

1. Mc Dowell, 1977, 45-46 ; j'ai analysé ces problèmes dans Engel, 1985 et Engel, 1989, chap. VIII.

(A) « Hespéros » dénote Hespéros
(B) « Phosphoros » dénote Phosphoros

qui décrivent la contribution faite par chaque nom aux conditions de vérité des phrases où ils sont susceptibles de figurer. Mais comme Hespéros est identique à Phosphoros, les côtés droits de ces biconditionnels peuvent être échangés. Mais il ne s'ensuit pas, de la vérité des phrases qui résulteraient de ces échanges, que les phrases échangées pourraient aussi bien faire office de théorie du sens, c'est-à-dire, pour satisfaire la condition de Frege, que quelqu'un qui connaîtrait les contenus des axiomes (A) et (B) connaîtrait pour autant les contenus des axiomes

(A') « Hespéros » dénote Phophoros
(B') « Phosphoros » dénote Hespéros

Il y aurait en tout cas une différence importante — précisément la différence en « valeur de connaissance » dont parle Frege — entre une théorie de la vérité qui utiliserait les spécifications (A) et (B) plutôt que les spécifications (B) et (C), qui se révélerait dans les théorèmes assignant des conditions de vérité aux phrases. Supposons donc que l'on formule une théorie répondant aux axiomes (A) et (B) pour « donner la signification » des noms en question dans le langage d'un locuteur. S'il se révélait que le locuteur est prêt à donner son assentiment à (A') et (B') également, alors la théorie-T qui aurait (A) et (B) comme axiomes serait aussi adéquate, pour saisir « la signification » des phrases du locuteur, que la théorie qui utiliserait les axiomes (A') et (B'). Mais si le locuteur ne donnait pas son assentiment à des phrases qui seraient des conséquences des axiomes (A') et (B') tout en donnant son assentiment à des phrases dérivées de (A) et (B), alors une théorie utilisant les axiomes (A') et (B') se révélerait inadéquate. C'est, pour l'essentiel, la procédure qu'utilise Davidson dans sa conception de l'interprétation. Cette procédure nous permet de comparer des assignations de conditions de vérité et des théories, mais elle ne nous dit pas directement ce qu'est le sens des expressions, pas plus qu'elle ne postule, comme le faisait Frege pour résoudre ce problème, l'existence de sens comme entités indépendantes. C'est précisément ce genre de décalages entre les théories extensionnellement équi-

valentes mais distinctes de la vérité qui est en cause dans les « contre-exemples » comme ceux de Foster à la thèse initiale de Davidson. Mais ces décalages ne menacent en rien la thèse véritable de Davidson, qui soutient qu'une procédure d'interprétation radicale doit nous permettre, dans certaines conditions empiriques, de dire quelle est, pour un locuteur ou une communauté de locuteurs données, la « bonne » théorie. Comme le dit McDowell, « la thèse [de Davidson] devrait être non pas que le sens est ce dont une théorie de la vérité est la théorie, mais plutôt que la vérité est ce dont une théorie du sens est la théorie ». Si l'on voulait utiliser la terminologie wittgensteinienne, on pourrait dire qu'une théorie de la vérité ne nous dit pas ce qu'est la signification, mais qu'elle nous le montre, à cette nuance près que ce qui se révèle ainsi n'a rien d'inexprimable. Tout cela est parfaitement conforme à la démarche indirecte de Davidson : caractériser — mais non pas définir ou analyser — la signification dans une langue naturelle à partir d'une caractérisation de la vérité, sans faire appel à des notions comme celles de signification, de manière à retrouver l'effet de telles notions :

> Il suffit peut-être, me semble-t-il, d'exiger que les phrases-T soient vraies. A l'évidence, cela suffit uniquement et correctement à déterminer l'extension du prédicat de vérité. Si nous considérons n'importe quelle phrase-T, cette proposition exige seulement que si une phrase vraie est décrite comme vraie, alors ses conditions de vérité sont données par une phrase vraie quelconque. Mais quand nous considérons le besoin contraignant qu'il y a à faire correspondre la vérité avec la vérité à travers tout le langage, nous nous rendons compte que n'importe quelle théorie acceptable selon cette norme peut effectivement produire un manuel de traduction utilisable qui permette de passer du langage-objet au métalangage. L'effet escompté est courant lorsqu'on construit des théories : extraire un concept riche (ici quelque chose de raisonnablement proche de la traduction) à partir de minces données fragmentaires (ici les valeurs de vérité des phrases) en imposant une structure formelle à un assez grand nombre de fragments. Si nous caractérisons les phrases-T par leur seule forme, comme le fit Tarski, il est possible, en employant les méthodes de Tarski, de définir la vérité en n'employant aucun concept sémantique. Si nous estimons que les phrases-T sont vérifiables, alors une théorie de la vérité montre comment nous pouvons passer de la vérité à quelque chose comme la signification — à quelque chose qui ressemble assez à la signification pour que

> si quelqu'un disposait d'une théorie pour un langage qui soit vérifié ainsi que je le propose, il pourrait utiliser ce langage dans la communication (1973 : 73-74, *119*).

C'est précisément ce qu'une théorie de l'interprétation doit établir.

2

Interprétation radicale

> When you come tomorrow, bring my football boots. Also, if humanly possible, Irish water spaniel. Urgent. Regards. Tuppy
> « What do you make of that, Jeeves ? »
> « As I interpret the document, Sir, Mr Glossop wishes you, when you come tomorrow, to bring his football boots. Also, if humanly possible, an Irish water spaniel. He hints that the matter is urgent, and sends his regards. »
> « Yes, that's how I read it too. »
>
> P. G. Wodhouse, *The Ordeal of Young Tuppy.*

2.1. Le problème de l'interprétation

La question initiale de Davidson : « Quelle forme devrait prendre une théorie de la signification pour une langue naturelle ? » ne s'identifie pas à la question : « Qu'est-ce qu'une TS (une sémantique) pour *L* ? », pas plus que sa réponse à cette dernière question (une TS doit prendre la forme d'une théorie-T) n'est suffisante pour répondre à la première. Elle ne suffit pas, comme on l'a vu, parce que la possibilité même d'appliquer une théorie-T à une *langue naturelle* suppose que certaines conditions *empiriques* soient réunies, outre les conditions formelles examinées au chapitre précédent. Ce problème se pose pour une langue naturelle et non pas pour une langue formelle, parce que les langues naturelles sont des instruments de communication. La question n'est pas, dans cette perspective, seulement celle de savoir ce que les mots et les phrases d'une langue naturelle signifient, indépendamment des locuteurs et des contextes, mais de savoir ce qu'ils signifient dans un certain usage et dans une certaine occasion. En ce sens, une théorie des conditions empiriques d'application d'une TS peut être considérée comme une théorie *pragmatique* de

la signification. Mais ce n'est pas la terminologie employée par Davidson, non parce que, comme on le dit souvent, il ignorerait les problèmes spécifiques posés par une pragmatique, mais parce qu'il entend poser ces problèmes au sein d'un cadre plus large, qui est précisément celui d'une théorie de l'*interprétation*. Telle que Davidson la conçoit, cette théorie comporte deux dimensions étroitement liées, l'interprétation du langage et l'interprétation des contenus mentaux et des actions.

La théorie de l'interprétation du langage concerne la question de savoir comment on peut établir, sur la base de ce qu'énonce un locuteur dans une occasion donnée, ce que ses mots signifient, et dans la mesure où il appartient à une communauté linguistique, comment il peut partager le langage de cette communauté. Il ne s'agit donc pas seulement d'établir une relation entre des expressions linguistiques et leur signification, mais aussi entre ces dernières et les locuteurs qui les utilisent. Le problème de l'interprétation est cependant étroitement lié à celui de l'établissement d'une TS pour une langue naturelle : interpréter ce qu'un locuteur signifie en une occasion donnée implique une connaissance de ce que ses mots signifient, ou pourraient signifier, en d'autres occasions, et par conséquent de ce que les expressions de son langage signifient. En ce sens, la connaissance d'une TS pour le langage d'un locuteur fait partie du savoir requis pour l'interprétation de ses énonciations. Il y a un lien direct entre le savoir requis par un interprète et la compétence sémantique du locuteur, telle qu'est supposée la spécifier une TS. Selon Davidson, *toute compréhension d'un langage suppose une capacité à l'interpréter*. En d'autres termes, si l'interprète d'un langage a besoin d'une compétence sémantique, cette compétence pourra être décrite comme une compétence de l'interprète. Ici encore, la question posée n'est pas celle des connaissances effectives que nous mettons en jeu quand nous interprétons le discours d'autrui, mais la question *normative* de savoir ce que nous *devrions* connaître pour interpréter.

Mais cela ne nous dit pas pourquoi il y a un problème de l'interprétation du langage. Supposons que nous ayons une TS pour un *L* que nous comprenons déjà. En ce cas, il nous suffira d'examiner les phrases-T produites par TS, et de voir si elles sont vraies. Il n'y aura pas de problème parce que nous pourrons vérifier si les phrases de TS (notre métalangage)

sont vraies quand celles de *L* (le langage-objet) le sont, puisque TS sera homophonique. Mais supposons que nous ayons affaire à un *L* que nous ne comprenons pas, dont les phrases ne sont pas interprétées. Dans ce cas, il ne suffit pas que les phrases-T de TS et celles de *L* soient homophones. Il faut encore que chaque phrase de *L* signifie la même chose que son homophone dans TS. Or c'est précisément ce que nous ne savons pas dans ce cas. Si nous voulons savoir si une théorie-T particulière s'applique à un *L* particulier, nous ne pouvons pas présupposer que les phrases de notre métalangage signifient la même chose que celles de *L*, sous peine de faire une pétition de principe. C'est relativement à ce type de situation, que Davidson appelle « interprétation radicale », que le problème de l'interprétation se pose : il s'agit de se placer dans la situation où nous ne connaissons pas le langage de ceux que nous avons à interpréter, et où nous ne disposons pas d'autres données que les énonciations et le comportement des locuteurs.

Le problème de l'interprétation se pose également dans le cadre d'une théorie de la pensée et de l'action. Nous n'interprétons pas seulement les contenus ou les significations d'expressions linguistiques, mais aussi des contenus de pensée, en attribuant à autrui et à nous-mêmes des attitudes propositionnelles, telles que des croyances, de désirs, de souhaits, ou des espoirs. Il est devenu courant, dans la philosophie contemporaine, d'appeler « psychologie populaire » le schème ordinaire d'explication de l'action et du comportement que nous utilisons pour interpréter nos semblables[1]. La question de l'interprétation dans ce domaine est la suivante : sur la base de quels principes et de quelles données pouvons-nous déterminer les contenus des attitudes postulées par la psychologie populaire ? Davidson fournit une réponse à cette question dans le cadre de sa philosophie de l'esprit et de l'action. Et la question revêt ici la même forme que celle qu'elle revêt dans le cadre de la théorie du langage. Supposons que nous ayons, dans une circonstance donnée, à interpréter une action particulière, par exemple l'action pour un individu d'allumer la lumière. Cette action ne peut être décrite comme telle — c'est-à-dire comme une

1. Cf. en particulier Fodor, 1981, 1987, et, pour une présentation de ces problèmes, Engel, 1988, 1992.

action d'*allumer la lumière* que si nous attribuons à l'agent certains états psychologiques, tels que le désir d'*allumer la lumière*, la croyance qu'*il allumera la lumière s'il tourne l'interrupteur*, ou l'intention d'*allumer la lumière*. Si nous n'attribuons pas ces états et leurs contenus psychologiques à l'agent, la seule chose que nous pourrons rapporter est un certain comportement observable, par exemple celui de tourner l'interrupteur, ou de mouvoir la main. Mais comment décrire l'action de l'agent comme étant celle d'*allumer la lumière* si nous ne connaissons pas ses états psychologiques ? Et comment déterminer le contenu de ces états si nous ne connaissons que son comportement observable ? Le problème est exactement similaire à celui qui consiste à assigner une certaine signification p à un énoncé s : si nous ne savons pas ce que signifie s pour le locuteur, nous ne pourrons pas déterminer s'il a énoncé que p, ou si s a le contenu de signification p. Et si nous ne savons pas quelles croyances a celui qui énonce s, nous ne pourrons pas non plus déterminer sa signification. Comme dans le cas du langage, nous ne pouvons pas présupposer, sans pétition de principe, ce qu'un agent croit, désire, ou veut, quand nous interprétons ses actions.

Mais le problème de l'interprétation linguistique des énoncés et celui de l'interprétation des pensées et des actions sont plus que similaires : ils sont étroitement liés. La plupart du temps, nous ne savons pas ce qu'un agent fait, quelles actions il accomplit, ni quelles croyances et désirs il a, si nous ne sommes pas en mesure d'interpréter ce qu'il *dit*, et si nous ne savons pas ce qu'un agent dit, si nous sommes pas en mesure d'interpréter ce qu'il *fait*, *désire* et *croit*. Appelons ce principe celui de *l'interdépendance des croyances et des significations*. C'est l'un des principes du *holisme* de Davidson. Ce n'est pas seulement un principe méthodologique, mais aussi et avant tout un principe substantiel, portant sur la nature même des croyances et des significations. Nous aurons à y revenir. Mais pour le moment on considérera son sens méthodologique. Une théorie de l'interprétation peut donc être représentée comme un triangle dont les trois sommets sont le langage, la pensée et l'action ; on ne peut déterminer l'un des facteurs sans déterminer les autres. Et, dans chaque cas, on ne peut supposer fixé l'un des facteurs sans considérer le problème comme résolu d'avance. La méthodologie est la même que celle qui gouverne le projet d'une théorie de la signification : il s'agit d'éviter de « réintro-

duire subrepticement dans les fondements mêmes de notre théorie des concepts qui seraient trop étroitement liés au concept de signification » (1984, xiii, 9). L'ensemble du projet de Davidson s'inscrit dans cette triple perspective, et c'est pourquoi ce qu'il vise n'est pas seulement une théorie du langage, mais « une théorie généralisée du langage et de l'action » (1980a) qui met en jeu ces trois dimensions. L'interprétation n'est pas seulement un processus qui intervient entre un interprète individuel et celui qu'il interprète, mais aussi un processus collectif mettant en jeu une communauté d'interprètes. En ce sens les significations, les pensées et les actions sont nécessairement *publiques*.

2.2. *Traduction radicale et interprétation radicale*

Le projet de Davidson s'inscrit dans la continuité de la conception quinienne de la « traduction radicale » (Quine, 1960). Mais il s'en distingue aussi de façon essentielle. On comparera les deux projets sur les quatre points suivants : 1 / quelle est la forme de la procédure envisagée ? 2 / quelles sont les données sur lesquelles elle se fonde ? 3 / comment ces données sont-elles organisées ? et 4 / quelles conséquences peut-on en tirer ?

1 / Quine aborde le problème de la signification à travers celui de la traduction, sans présupposer la notion de signification, ou les notions voisines de synonymie et d'analyticité. Il pose la question de savoir comment il est possible d'établir un manuel de traduction entre deux langues L et L'. Ce manuel consistera en une fonction, appliquée récursivement, qui pour toute phrase de L donnera une phrase de L'. On veut savoir quelles sortes de règles doivent gouverner une traduction correcte, à partir des données empiriques minimales. Le seul moyen d'envisager le problème sous sa forme pure, c'est-à-dire sans introduire des données qui présupposeraient que des règles de traduction aient déjà été établies, est de considérer la situation de traduction *radicale* d'une langue encore inconnue, pour laquelle il n'existe aucun manuel de traduction disponible. A aucun moment, il n'est supposé qu'il s'agisse d'autre chose que d'une expérience de pensée.

2 / Quine présuppose que les données dont disposera le traducteur radical sont des données purement physiques. Ce seront des descriptions de l'environnement des locuteurs et des séquences causales de cet environnement, des descriptions des stimulations sensorielles qu'ils reçoivent et des émissions linguistiques concrètes qu'ils produisent. Quine s'en tient à ces données physiques d'une part parce que l'emploi d'autres données — psychologiques ou sémantiques — présupposerait ce qui est en question, et d'autre part parce que son objectif est de reconstruire la notion de signification à partir des seules données physiques et comportementales, dans un cadre béhavioriste.

3 / Un simple flux d'événements physiques ne suffit pas pour engager le processus de traduction. Pour prendre pied dans un manuel de traduction, le traducteur radical doit collecter un stock d'énoncés de l'indigène prononcés dans un certain environnement en présence de certaines stimulations, les transformer en questions posées à l'indigène, et noter ses manifestations d'assentiment et de dissentiment. Ce sont ces manifestations qui serviront de base au processus de traduction. Par généralisation à partir de ces situations initiales, on construit une corrélation entre les structures de stimulations sensorielles et les phrases prononcées en ces occasions. Cela permet de dégager la « stimulation stimulus » (« *s* - signification »), affirmative ou négative d'une phrase donnée, comme la classe de toutes les stimulations qui provoquent l'assentiment ou le dissentiment de l'indigène. A chaque type de *s* - signification ainsi identifiée, le linguiste « de jungle » associe des hypothèses de traduction, ou « hypothèses analytiques », qui lui permettent d'établir progressivement son manuel. On peut alors établir un ordre de complexité entre les phrases, selon la plus ou moins grande ressemblance des *s* - significations qui leur sont associées : les phrases « occasionnelles » (« Oh ! un lapin ! »), qui sont les produits de réponses directes à ces stimulations, les phrases « d'observation » (« Ceci est un lapin »), qui dépendent d'un apprentissage préalable et d'une stimulation présente, et qui varient selon l'information collatérale, puis des phrases contenant des termes identifiés comme étant plus ou moins théoriques (« Ceci est un animal »). Cette distinction n'est cependant que de degré, puisque Quine refuse une séparation forte entre termes observationnels et termes théoriques. Quine entend reconstruire les relations de

signification entre phrases, à partir des notions de « *s* - analyticité » (propriété des phrases pour lesquelles le sujet donne son assentiment sous toutes les stimulations) et de « *s* - synonymie » (couples de phrases telles que toutes les stimulations provoquent l'assentiment aux deux phrases en même temps).

4 / Quine soutient que l'on peut ainsi reconstruire un manuel de traduction établissant une corrélation systématique entre les phrases du locuteur indigène et le langage du traducteur, et qu'il n'y a pas d'autre manière raisonnable de construire un tel manuel. Le problème est : ce manuel sera-t-il correct ? Et la réponse célèbre de Quine est : non. Toutes les données que peut recueillir le traducteur radical, tous les faits qu'il peut établir quant aux corrélations entre des structures de stimulations et des phrases sont insuffisantes pour déterminer un schème de traduction unique du langage indigène dans celui du linguiste. Une pluralité de manuels pourront être construits, incompatibles entre eux, bien que compatibles avec la totalité des dispositions des locuteurs et des données comportementales et physiques. C'est la formulation la plus courante de la thèse de « l'indétermination de la traduction » (*IT*) . Elle revient, sous cette forme, à affirmer la *sous-détermination* d'un manuel de traduction par les données empiriques du linguiste. Mais cette sous-détermination n'implique pas une *indétermination* de toutes les hypothèses analytiques du traducteur. En particulier Quine admet que la traduction des connecteurs logiques qui sont des fonctions de vérité (« et », « non », « ou », etc.) peut reposer sur des hypothèses plus fermes que pour les autres expressions. Nous ne pouvons pas, en traduisant un langage indigène, manquer de préserver les lois logiques qui gouvernent notre propre langage, en employant le principe que Neil Wilson (1959) appelle le « principe de charité », consistant à « maximiser » le nombre de vérités logiques que nous pouvons découvrir dans le langage indigène. Mais il n'en est pas de même pour la traduction des termes référentiels, tels que les noms et les quantificateurs. Dans ce cas, il n'y a pas moyen de dire, sur la base des données comportementales, à quoi les termes singuliers ou les prédicats font référence, parce qu'il y a un grand nombre de schèmes de référence possibles compatibles avec les données. C'est le point qu'illustre le célèbre exemple de l'énoncé « Gavagai », qui peut signifier qu'il y a ici un

lapin, de la lapinité, des parties temporelles coprésentes de lapin, etc. La référence, singulière ou générale, est « inscrutable », et c'est ce qui constitue la source principale de l'indétermination de la traduction et de la signification. A cause d'elle, l'indétermination affecte toute traduction : aucune traduction ne peut jamais être correcte, y compris la traduction homophonique des énoncés de ceux qui parlent notre propre langue : l'indétermination commence « chez nous ». La conséquence, pour la notion de signification, dont la procédure du manuel de traduction proposait une reconstruction, est immédiate : il n'y a pas, relativement aux données observables, de *fact of the matter*, de fait décisif, permettant d'établir la signification d'une expression d'un langage quelconque.

Quine a cependant un autre argument que celui de la traduction radicale en faveur de IT, fondé sur le holisme de la signification et sur le holisme épistémologique[1]. Cet argument part de deux prémisses : *a*) le principe de la théorie vérificationniste ou empiriste de la signification : la signification d'un énoncé est la différence que cet énoncé fait par rapport à la totalité de l'expérience possible, et *b*) le principe du holisme épistémologique : il n'y a pas *une* seule différence que fait un énoncé par rapport à l'expérience, mais *l'ensemble* des énoncés sont vérifiés par rapport à la totalité de l'expérience possible, et conclut *c*) que la signification d'un énoncé est nécessairement indéterminée[2].

Le trait le plus frappant de la conception quinienne de la signification est l'étroitesse du cadre empirique et des présupposés qu'elle met en œuvre — le béhaviorisme et le physicalisme — en sorte qu'on se demande souvent si sa conclusion n'est pas simplement une conséquence de ces prémisses très restrictives. On peut se demander si en changeant les prémisses de l'expérience de pensée et la nature de la procédure on peut parvenir à des conclusions similaires. La conception davidsonienne de l'interprétation radicale est précisément l'une des manières de répondre à cette question.

1. Le premier argument, celui de Quine, 1960, chap. 2, est celui qu'il appelle l'« argument d'en bas » [*from below*] ; le second, celui du holisme, est celui qu'il appelle « d'en haut » [*from above*] dans Quine, 1970.

2. C'est, pour l'essentiel, la fameuse image de la science comme un champ d'énoncés rencontrant l'expérience à sa périphérie, proposée dans « Two Dogmas of Empiricism » (Quine, 1951). Ce qui précède n'a pas pour but de donner autre chose qu'une présentation sommaire des thèses quiniennes.

Comme Quine, Davidson entend reconstruire les conditions générales qui gouvernent toute compréhension d'un langage à partir d'une situation hypothétique « radicale » au sens indiqué ci-dessus. Mais il change profondément la nature et les conditions du problème posé par Quine. Reprenons les questions 1 /-4 / dans le même ordre que précédemment.

1 / Pour Quine comme pour Davidson, l'objet d'une traduction ou d'une interprétation d'un langage est la détermination des significations des énoncés d'un locuteur et d'une communauté donnée. Pour Quine, la procédure de traduction radicale fait l'économie de l'idée que les locuteurs ont des états psychologiques, puisque les significations sont seulement les corrélats de structures de stimulation sensorielles et de dispositions comportementales. Deux locuteurs ne peuvent être dits partager le même langage et attribuer les mêmes significations à des expressions que s'ils partagent les mêmes structures et dispositions. Cela découle de la prémisse physicaliste et béhavioriste de Quine, pour qui les états intentionnels des locuteurs n'ont aucune réalité en dehors des dispositions comportementales qui leur correspondent. Pour Davidson au contraire, on ne peut pas interpréter des significations sans interpréter des contenus d'attitudes propositionnelles des locuteurs. De plus ceux-ci ne peuvent être définis en termes béhavioristes. Cela suppose que l'on donne aux états intentionnels une réalité beaucoup plus substantielle que celle d'être seulement des structures dispositionnelles. La réalité minimale que Davidson confère à ces états est d'être des coordonnées indispensables dans la procédure qui consiste à attribuer des significations : comme on l'a vu, on ne peut attribuer une signification à un locuteur sans lui attribuer en même temps certains contenus de croyances, de désirs ou d'intentions, ni sans interpréter ses actions. Or l'une des thèses fondamentales de la théorie davidsonienne de l'action (1980) est qu'une action est un certain type d'événement, qui peut être décrit sous un aspect qui le révèle comme étant le produit d'une certaine attitude mentale qui en est la raison ou la cause. Tout comme les mouvements corporels d'un agent, les énonciations d'un locuteur sont des actions, qui peuvent être décrites comme le produit d'attitudes mentales. Assigner, par conséquent, une signification à un énoncé au moyen d'une phrase de la forme « *X* dit que *p* », c'est décrire l'événement constitué par une énonciation d'une phrase *s* d'une

certaine manière, et cet événement peut être redécrit au moyen d'une description psychologique de la forme « *X* croit que *p* ». Il s'agit là aussi de l'une des hypothèses majeures qui sous-tendent la théorie davidsonienne du discours indirect (§ 1.4). Le problème est de savoir comment on peut relier ces descriptions à l'événement qu'est l'énonciation, et relier les descriptions entre elles (1974 ; 1975 : 161, *236*). La nécessité de passer par l'« idiome intentionnel » est l'une des raisons pour lesquelles Davidson parle d'*interprétation* plutôt que de *traduction*.

L'autre raison de ce changement de terminologie est que Davidson n'entend pas établir un manuel de traduction. Comme on l'a vu (§ 1.3), une TS pour *L* ne prend pas la forme d'une théorie de la traduction entre deux langages, mais porte sur un seul langage. Certes le langage (métalangage) que nous utilisons pour énoncer la théorie est distinct du langage-objet qui est le sujet de la théorie, mais il n'en est pas l'objet propre. En revanche, une théorie de la traduction implique trois langages : le langage-objet (à traduire), le langage-sujet (dans lequel le langage-objet est traduit), et le métalangage (le manuel de traduction). Mais on a vu qu'il est possible de connaître un manuel de traduction sans savoir ce que les phrases du langage-objet et celles du langage-sujet signifient. Un manuel de traduction ne peut donc se substituer à une théorie de l'interprétation, et même s'il peut s'y ajouter, il est inutile, à partir du moment où nous essayons de donner directement la théorie d'un langage-objet dans un langage qui nous est déjà connu.

Quelle forme prendra une théorie de l'interprétation ? Nous connaissons déjà la réponse : elle prendra la forme d'une théorie-T, c'est-à-dire d'une théorie capable de révéler la structure du langage-objet, et satisfaisant les conditions formelles adéquates pour être adaptée à une langue naturelle. Cette réponse suffit à distinguer le projet de Davidson de celui de Quine : ce dernier ne met l'accent ni sur la notion de vérité, ni sur l'idée qu'un manuel de traduction doit discerner une structure dans le langage-objet (1977a : 205, *298*). Néammoins, même si une théorie de l'interprétation n'est pas un manuel de traduction, on exigera d'elle qu'elle puisse nous fournir l'équivalent de ce que l'on peut tenir comme une traduction correcte des phrases d'un *L*. Pour anticiper la réponse de Davidson à la question 4 /, il y a aura pour lui, contrairement à Quine, quelque chose comme une traduction *correcte*.

2 / Pour Davidson comme pour Quine, les données sur lesquelles s'appuie une théorie de l'interprétation doivent être essentiellement non sémantiques (le sens de cette restriction « essentiellement » pour Davidson apparaîtra plus loin). On vient de voir que pour Davidson ces données ne peuvent pas être purement comportementales. Elles ne sont pas, en particulier, les conditions de stimulations sensorielles qui, selon Quine, « déclenchent » l'assentiment ou le dissentiment face à une phrase, et qui constituent les raisons empiriques qu'a un locuteur de donner son assentiment ou son dissentiment. En fait Davidson rejette l'empirisme quinien qui suppose que soient associées à chaque type de phrase des données sensorielles *proximales* qui la vérifient. Selon lui, les significations des phrases du langage d'un individu doivent être rattachées aux objets et événements *distaux* qui figurent dans l'environnement commun à l'interprété et à son interprète[1]. L'interprète suppose donc que ce sont les objets saillants dans cet environnement sur lesquels portent les phrases de l'interprété, sans recourir à l'idée que des stimulations sensorielles confirment ces phrases. Ce sont donc les objets du monde extérieur, tels qu'ils sont donnés intersubjectivement, qui servent à l'interprète à établir la référence des phrases. L'interprète va supposer que l'interprété entre dans une relation causale avec ces objets, mais il n'y a aucune raison de postuler des intermédiaires entre le sujet et ces objets, comme des stimulations sensorielles. Cela veut dire que Davidson rejette la notion quinienne de stimulus-signification (1979 ; 1990b). Et c'est pourquoi il est trompeur de parler ici de « données » si cela doit suggérer que Davidson souscrit à une théorie vérificationniste de la signification.

L'autre type de « données » qu'envisage Davidson est psychologique : ce sont les croyances et autres attitudes des agents. Mais ce sont des données que, à l'exception des croyances que l'interprète s'attribue à lui-même, il ne peut connaître qu'indirectement chez les sujets, à travers ce qu'ils disent et font. Le problème de l'interdépendance des croyances, des significations et des actions a un analogue en théorie de la

1. Les stimuli proximaux, dans la terminologie béhavioriste, sont ceux qui affectent causalement les récepteurs sensoriels de l'individu, les stimuli distaux sont ceux qui partent des objets et événements de l'environnement extérieur au sujet. Sur la théorie distale de la référence, cf. ci-dessous, § 6.3.

décision[1]. En théorie de la décision, le choix d'une action plutôt qu'une autre, ou la préférence qu'un événement ou état de choses ait lieu plutôt qu'un autre est le produit de deux facteurs : la valeur (ou la désirabilité ou l'utilité) que l'agent assigne à une action possible, et les croyances de l'agent concernant la probabilité des conséquences de cette action. On suppose que ces facteurs sont mesurables : qu'il y a des degrés de croyances, ou probabilités subjectives, et que les désirs ou valeurs ont également une cardinalité. On suppose que l'agent, en choisissant des actions possibles (généralement représentées par des paris), en choisit une dont les conséquences ont la valeur relative la plus grande, c'est-à-dire « maximise son utilité espérée », et qu'il est, en ce sens, rationnel. Un trait important de cette théorie est qu'elle distingue quelque chose d'observable — des préférences, des actions, ou des choix — de quelque chose qui ne l'est pas, et qui a un statut relativement théorique : des degrés de croyances et de désirs. Ramsey (1926) a proposé une procédure expérimentale permettant de calculer le degré de croyance d'un agent en une certaine proposition, une fois données ses préférences et ses valeurs, ou inversement de calculer ses valeurs, étant donné ses degrés de croyance et ses préférences. Mais étant donné seulement ses préférences ou ses choix, comment calculer ses valeurs et ses croyances ? Ramsey résout le problème en montrant qu'on peut trouver une proposition considérée comme aussi probable que sa négation, sur la base de choix simples, puis qu'on peut utiliser cette proposition pour construire une série infinie de choix de paris parmi lesquels on peut déterminer une mesure de la valeur de toutes les options et de leurs probabilités[2]. Le parallélisme avec la théorie de l'interprétation est clair. Aux degrés de croyances et aux désirs postulés par la théorie de la décision correspondent les croyances et les significations en théorie de l'interprétation. La partie observable est constituée dans un cas par les préférences ou les choix, et dans l'autre par le comportement verbal. Dans chacun des cas, on ne peut déterminer un élément sans déterminer les deux autres. Si l'on poursuit l'analogie, ce dont nous avons besoin en théorie de l'interprétation est une donnée qui

1. Davidson s'appuie ici sur la théorie de Ramsey, qu'il a développée dans ses travaux sur l'utilité (Davidson, Suppes & Siegel, 1957). Cf. Davidson, 1974a, 1975, 1976a, 1980, 1985, 1990.
2. Pour un exemple, cf. 1974a.

soit l'homologue de la notion de préférence ordinale entre des paris qui sert de levier pour déterminer les degrés de croyance et les différences de valeurs. Davidson propose que cette donnée soit une certaine attitude propositionnelle, celle de *tenir-pour-vraie* (*holding true*) une certaine phrase. Il s'agit d'une attitude propositionnelle car c'est une *croyance* qu'une phrase *s* est vraie. Mais elle peut être isolée par l'interprète sans que celui-ci sache quelle croyance elle exprime, et sans qu'il sache non plus quelle est sa signification. En ce sens, tenir-pour-vrai « *p* » n'est pas la même chose que *juger que p*, ou qu'accepter que *p*, qui supposent que le sujet connaisse le contenu ou la signification de *p*. Supposons, par exemple, que je lise dans un ouvrage scientifique que « la sérotonine est dérivée du tryptophane ». Je peux savoir que cette phrase est vraie sans en connaître la signification, ni sans savoir *quelle* vérité elle exprime. De la même manière, Davidson suppose que l'interprète radical peut identifier la présence de cette attitude chez un agent, tout comme on peut observer les préférences en théorie de la décision. Il pourrait sembler que la notion de tenir-pour-vrai ne soit qu'une transposition de la notion quinienne d'assentiment à une phrase provoqué par des stimulations. Mais s'il s'accorde avec Quine sur le fait que la phrase soit l'unité de base de l'interprétation linguistique, Davidson insiste sur le fait que la notion de tenir-pour-vrai est une notion *intentionnelle*, qui n'est pas réductible à d'autres notions plus primitives comportementales. C'est ici que la restriction (« essentiellement ») mentionnée plus haut prend son sens : il est faux de dire que Davidson ne s'appuie sur *aucun* fait sémantique ou intentionnel dans sa théorie de l'interprétation et de la signification. Le fait qu'un locuteur tienne une phrase *s* pour vraie est bien un fait sémantique, mais c'est un fait sémantique *minimal*, qui ne présuppose pas des contenus d'attitudes ou des faits sémantiques plus riches.

2.3. Le principe de charité

Venons-en maintenant à la question 3 / : comment les données dont dispose l'interprète peuvent-elles être utilisées pour construire une théorie-T pour le langage étudié ? L'interprète collecte des données concernant les

phrases tenues pour vraies dans certaines circonstances, par exemple pour un locuteur allemand, Kurt :

(1) Kurt appartient à la communauté linguistique allemande et Kurt tient pour vrai « Es regnet » samedi à midi, et il pleut près de Kurt samedi à midi.

L'interprète émet, sous la forme d'une phrase-T, l'hypothèse que :

(2) « Es regnet » est vrai en allemand si énoncé par x au temps t ssi il pleut près de x au temps t.

Rappelons que les phrases-T doivent être relativisées à un lieu, un temps, et un locuteur. La mention de la « communauté linguistique allemande » n'est pas une pétition de principe, car deux locuteurs appartiennent à la même communauté linguistique s'ils peuvent être soumis à la même théorie de l'interprétation (1973, 135, *187*). Une première contrainte de l'interprétation est que l'interprète ne doit pas se contenter de faire des hypothèses comme (2) sur la base d'énonciations isolées du ou des locuteurs : il ne peut les faire que sur un ensemble d'énonciations récurrentes, et sur un ensemble de circonstances. Cela suppose une forme d'induction à partir d'hypothèses isolées qui conduise à des généralisations du type suivant :

(3) (« $\forall x$) (« $\forall y$) (si x appartient à la communauté linguistique allemande alors x tient pour vrai « Es regnet » en t ssi il pleut près de x en t.

Le fait qu'il pleuve près du locuteur est la donnée (causale) qui permet à l'interprète de former son hypothèse. Cela suppose qu'il croie lui-même qu'il pleut, et par conséquent qu'il croie que la phrase « Il pleut près de Kurt samedi à midi » est vraie. L'interprète peut se tromper : il n'est pas supposé infaillible. En ce cas, s'il s'en aperçoit, il pourra réviser sa croyance. Mais il doit supposer que ses propres croyances sont vraies. Le locuteur aussi peut se tromper. Il n'est donc pas possible de tenir les phrases-T comme (2) et (3) comme vraies au sens où elles seraient d'emblée *établies* comme vraies. Mais l'interprète doit, au moins dans la première étape de son interprétation, les tenir pour *généralement* ou *de prime abord* vraies. C'est ici qu'intervient le principe qui constitue la clef de voûte de la méthode d'interprétation de Davidson, le *principe de charité* (PC). Il a plusieurs sens.

En premier lieu, le PC est une *règle méthodologique*, ou une maxime de l'interprétation. Quine, comme on l'a vu, use du PC comme d'une condition de traduction des constantes logiques, et l'emploie comme une maxime destinée à écarter de la part du traducteur toute imputation d'erreurs par violation de principes logiques. Chez Quine, le PC revient à dire que toute traduction doit préserver les lois logiques de la logique usuelle (classique). C'est en ce sens qu'il dit que « la stupidité de notre interlocuteur est, au-delà d'un certain point, moins probable qu'une mauvaise traduction » et que « la mentalité prélogique est un trait infusé par de mauvais traducteurs » (1960 : 59)[1]. En ce sens, le PC est un principe de *cohérence logique* ou de *rationalité* des croyances, qui nous prescrit de tenir celles-ci comme étant, dans l'ensemble, rationnelles, selon les canons de la logique. On peut le formuler ainsi :

(PCR) *Interprétez toujours de manière à rendre les croyances de ceux que vous interprétez cohérentes et non contradictoires.*

Mais Davidson ne limite pas le PC à cette maxime de rationalité : il l'étend, à la différence de Quine, à l'ensemble de la procédure interprétative, « systématiquement » (1973 : 136, *203*; 1984 : xvii, *15*). Dans ce cas, le PC prend la forme d'un principe de *véridicité* des croyances[2]. Il s'agit d'abord, pour l'interprète, de présupposer la vérité de ses propres croyances. Et il s'agit également pour lui de présupposer la vérité des croyances de l'interprété. La formulation privilégiée de Davidson, dans ses premiers écrits[3], est que le principe prescrit à l'interprète de « maximiser l'accord » et de « minimiser le désaccord » entre ses croyances propres et celles de l'interprété :

(PCV) *Interprétez toujours de manière à maximiser l'accord entre vos croyances et celles de ceux que vous interprétez.*

On peut comprendre pourquoi le PCV est indispensable à la méthode d'interprétation. Puisque l'interprète cherche à construire une théorie-T

1. J'ai analysé cette critique quinienne de Levy-Bruhl dans Engel, 1989a.

2. Dans 1991c : 158, Davidson propose d'appeler PCR « Principe de cohérence », et PCV ci-dessous « Principe de Correspondence ».

3. 1967 : 27 ; 1973 : 136-137 ; 1974 : 196-197 ; 1974a : 152-153 ; 1975 : 168-169.

pour le langage qu'il interprète, il doit, pour construire ses phrases-T, d'une part répérer les phrases que l'interprété tient pour vraies, et d'autre part forger ses hypothèses en fonction de ce qu'il tient lui-même pour vrai, dans *son* langage, qui sera le métalangage de sa théorie-T. Les phrases qu'il tient lui-même pour vraies, ses propres croyances, vont donc lui servir de point d'appui pour construire sa théorie. Sans l'hypothèse que les locuteurs qu'on interprète ont des croyances correctes pour la plupart, il est impossible de résoudre le problème de l'interdépendance de la croyance et de la signification. Le point d'appui va consister à « tenir la croyance comme constante autant que possible tout en cherchant à résoudre la signification » (1973 : 137, *203*). Toute la méthodologie « minimaliste » de Davidson pour la construction d'une TS repose donc sur l'idée que, bien que nous ne connaissions pas la signification des phrases du *L* interprété, nous devons supposer la vérité comme *déjà comprise*, ou comme partagée, par l'interprète et par l'interprété. Cela veut-il dire que l'interprète et l'interprété tiennent tous deux pour vraies les *mêmes* croyances, c'est-à-dire qu'ils ont *toutes* les vérités en commun, c'est-à-dire qu'ils sont infaillibles ? Evidemment pas, sans quoi l'interprétation serait toujours facile, et l'accord permanent. L'expérience la plus commune, quand on interprète, est l'expérience de ne pas comprendre. Mais Davidson suppose que, pour qu'on puisse avoir cette expérience, il faut aussi qu'il existe un fonds commun de compréhension. Le problème est qu'on ne sait pas exactement lequel. C'est pourquoi Davidson dit souvent que le PC prescrit que la *plupart*, ou un très grand nombre, de croyances de l'interprété sont vraies. Toute la méthodologie de l'interprétation repose sur l'idée qu'*il y a du vrai, bien qu'on ne sache pas exactement où il tombe*. Le but de l'interprétation est de parvenir à savoir où il tombe, si nous voulons aller de la vérité à la signification. En ce sens, le PC n'est qu'une conséquence de la méthodologie qui consiste à ne pas exiger autre chose, pour une TS, qu'une *théorie*, et non pas une *définition* de la vérité (§ 1.3.3) Sans ce principe, la méthode perd son sens.

Mais cela ne nous permet pas encore de comprendre comment fonctionne le PC dans la méthode d'interprétation. Si on le prend comme la conjonction de PCR et de PCV, le PC peut encore être lu en deux sens apparemment distincts. On peut tout d'abord le comprendre comme nous

prescrivant d'attribuer à autrui ce qu'il *serait* rationnel de croire, dans telles et telles circonstances, en un sens normatif : *efforcez-vous d'attribuer à autrui les croyances qu'il serait correct d'avoir*. Mais la question se pose de savoir selon quelles normes l'interprète doit faire ses attributions : s'agit-il de *ses* propres normes, de ce qu'il juge lui-même être rationnel, vrai, correct ? Mais dans ce cas, comment éviter des conflits éventuels entre les normes de l'interprète et celles de l'interprété ? Ou bien s'agit-il de normes transcendantes par rapport à l'interprète comme le sont, en principe, les lois de la logique ? Mais dans ce cas, quelles raisons avons-nous de croire que l'interprété suit lui-même ces normes ? On peut comprendre aussi le PC en un autre sens, comme prescrivant d'attribuer à l'interprété non pas ce qu'il serait, en général, et pour tout individu possible, rationnel et correct de croire, mais comme lui prescrivant d'attribuer ce qu'il est correct de croire, *pour lui interprète*. En ce cas, on se figure que l'interprète essaie d'attribuer à l'interprété le maximum de croyances *similaires* aux siennes, en projetant son univers doxastique et sa propre psychologie sur celui de l'interprété.

Le contraste entre les deux lectures apparaît clairement si l'on considère l'objection qui apparaît la plus évidente à l'usage d'un tel principe : un agent ne peut-il pas se tromper dans ses croyances, en tenant certaines phrases comme vraies, et ne peut-il pas faire des inférences incorrectes et être en ce sens irrationnel ? Dans ces cas d'erreur, le PC ne nous prescrit-il pas, contre toute évidence, de tenir néammoins cet individu pour vérace et rationnel ? Ne peut-on même pas supposer qu'un sujet soit « massivement » dans l'erreur ? Cependant l'erreur ou l'irrationalité ne paraissent faire problème que si l'on adopte la lecture normative du principe : si l'on adopte au contraire la stratégie consistant à projeter nos propres croyances sur l'univers de croyances de l'interprété, rien ne s'oppose à ce que nous comprenions ce dernier sur la base d'attributions de croyances que nous tenons nous-mêmes comme *fausses*. De nombreux auteurs ont soutenu en ce sens que le PC ne pouvait pas être une bonne maxime d'interprétation, et ont défendu l'idée que le seul principe effectif devait être un « principe d'humanité » d'après lequel nous devons attribuer aux sujets le maximum de croyances similaires aux nôtres, de manière à pouvoir les comprendre (Grandy, 1973, Mc Ginn, 1977, Goldman, 1989).

En fait Davidson lui-même semble souvent opter pour la seconde lecture. Il souligne que le véritable objectif du PC n'est pas tant de maximiser l'accord, c'est-à-dire d'attribuer à l'interprété le maximum de croyances vraies, que d'« optimiser » cet accord, c'est-à-dire de rendre aussi similaires que possible l'ensemble de croyances de l'interprète et de l'interprété, qu'elles soient vraies ou fausses (1975 ; 169, *249*). Il ne s'agit pas de rechercher à tout prix l'accord, mais de rechercher le *meilleur accord possible*. Mais ailleurs Davidson rejette aussi cet idéal :

> Minimiser le désaccord, ou maximiser l'accord, est un idéal confus. Le but de l'interprétation n'est pas l'accord, mais la compréhension. J'ai toujours insisté sur le fait que l'on ne peut parvenir à la compréhension qu'en interprétant de façon à obtenir la bonne espèce d'accord. Mais la « bonne espèce » d'accord n'est pas plus facile à spécifier que de dire ce en quoi consiste une bonne raison d'accepter une certaine croyance (1984 : xvii, *15*).

Selon cette ligne de pensée, le PC n'a pas pour but d'attribuer systématiquement à un sujet des croyances vraies ou rationnelles, et n'exclut en rien l'erreur de sa part. Il nous prescrit seulement de minimiser l'erreur *inexplicable* :

> Certains désaccords sont plus propres à détruire la compréhension que d'autres, et une théorie sophistiquée doit prendre ce point en considération. Le désaccord quant à des sujets théoriques peut (dans certains cas) être plus tolérable que le désaccord quant à ce qui est plus évident ; le désaccord quant à la manière dont les choses nous apparaissent est moins tolérable que le désaccord quant à ce qu'elles sont en réalité ; le désaccord quant à la vérité des attributions de certaines attitudes à un locuteur par ce même locuteur peut ne pas être tolérable, ou à peine tolérable. Il est impossible de simplifier les considérations pertinentes, car tout ce que nous savons où croyons quant à la manière dont nos croyances sont confirmées peut être mis à profit pour décider où la théorie peut le mieux s'accommoder avec l'existence d'erreurs, et pour décider quelles sont les erreurs qui détruisent le moins la compréhension. La méthodologie de l'interprétation n'est, de ce point de vue, rien d'autre que l'épistémologie au miroir de la signification (1975 : 169, *204*).

Mais si l'interprétation est « l'épistémologie au miroir de la signification », le PC n'est alors plus seulement une règle d'attribution à autrui de croyances vraies « pour la plupart » ou similaires aux nôtres. Il suppose que les

normes que nous acceptons pour justifier nos propres croyances conditionnent toute interprétation. *Tout ce que nous tenons comme correct, et nos critères mêmes de correction*, doivent intervenir. En ce sens, il n'y a aucun conflit entre la lecture « normative » du principe et sa lecture « psychologique ». Davidson veut souligner qu'il n'y a pas de différence précise entre ce que croit un interprète et ce qu'il croit que les autres devraient croire. Il projette aussi bien sa propre psychologie et ses croyances que les raisons qu'il peut avoir pour soutenir ses croyances, et il n'y a pas de conflit entre le PC et le principe « d'humanité ». La distinction entre le sens normatif du principe, qui nous prescrit d'idéaliser la véridicité et la rationalité des sujets, et son sens psychologique, est destinée à s'estomper, si nous nous rappelons que le PC, en tant que règle méthodologique, est supposé guider l'interprète dans les premières étapes de son interprétation, et non pas prescrire une assignation définitive de croyances vraies et rationnelles. Davidson insiste régulièrement sur le fait que le processus d'interprétation est un processus dynamique, au cours duquel l'interprète peut réviser ses attributions initiales, qui sont comme autant d'hypothèses que la suite confirmera ou infirmera.

Mais pourquoi devrait-on admettre le principe de charité ? N'est-ce pas une hypothèse extrêmement coûteuse et totalement irréaliste de supposer que la plupart des croyances agents sont vraies et rationnelles, alors même que l'erreur et l'irrationalité semblent au contraire la règle ? La plupart des critiques de Davidson sur ce point font remarquer que le PC manque totalement de plausibilité psychologique et qu'on ferait mieux, s'il doit servir de principe d'interprétation du comportement, adopter un principe plus réaliste (comme le principe d'humanité)[1]. Ce raisonnement repose sur deux prémisses. La première est que l'on pourrait *découvrir*, à titre de fait, que l'erreur et l'irrationalité sont la règle et non l'exception, dans les croyances et le comportement humain. La seconde est que l'on pourrait découvrir qu'elles sont *massives*. Mais les deux prémisses sont contestables. Car Davidson conteste précisément que l'on puisse découvrir empiriquement que les individus sont rationnels et non véraces. Or cela suppose que nous ayons déjà identifié leurs croyances et leurs

1. Cf. par exemple Nisbett et Thagard, 1983.

comportements. Mais identifier une croyance ou un comportement suppose déjà que l'on ait pu discerner chez l'individu un *minimum* de rationalité et de véracité. En d'autres termes, le PC n'est pas seulement une règle méthodologique, que nous pourrions décider de suivre ou de ne pas suivre, une maxime utile dont on pourrait comparer les avantages et les inconvénients par rapport à d'autres maximes possibles. Davidson soutient que le PC est un principe *normatif* et *a priori*, la condition de possibilité minimale de toute interprétation et de toute compréhension. Nous ne pourrions pas interpréter un individu quelconque si nous ne supposions qu'il est largement vérace et minimalement rationnel dans ses croyances. En fait nous ne pourrions même pas lui attribuer des croyances :

> Il ne faudrait pas croire que le conseil méthodologique d'interpréter, de manière à optimiser l'accord, repose sur un présupposé charitable concernant l'intelligence humaine qui puisse se révéler faux. Si nous ne pouvons pas trouver une manière d'interpréter les paroles et les autres comportements d'une créature comme révélant un ensemble de croyances largement cohérentes et vraies selon nos propres normes, nous n'avons aucune raison de considérer que cet être est rationnel, qu'il a des croyances, ou qu'il dit quoi que ce soit (1973 : 137, *203*).

En d'autres termes, il n'y a pas d'autre principe possible que le PC : « La charité nous est imposée : que nous le voulions ou non, si nous voulons comprendre les autres, nous devons considérer qu'ils ont raison sur la plupart des sujets » (1974 : 197, *287*). Sur quel argument repose cette assertion ? Elle ne repose pas seulement sur l'idée que la méthode d'interprétation doit supposer la vérité déjà comprise. Elle repose sur la thèse du holisme dont nous avons vu qu'elle était l'une des conditions constitutives d'une théorie de la signification selon Davidson. Mais il ne s'agit pas ici du holisme de la signification, mais du holisme de la *croyance*. Davidson insiste régulièrement sur ce point : un état mental n'est pas une croyance si nous ne pouvons pas le relier à d'autres croyances. Mais relier une croyance à d'autres croyances et à d'autres états mentaux, c'est découvrir une structure rationnelle minimale dans la créature, un ensemble quelconque de *raisons* qu'elle a de croire ce que nous lui attribuons, ou d'agir de telle façon. C'est précisément parce que nous ne pouvons pas être sûrs de la présence de cette structure rationnelle minimale chez les animaux que nous ne pouvons pas valablement leur attribuer des

pensées et des croyances, ni même un langage (1975, 1982, cf. *infra* § 6.5). Mais l'interdépendance ou la cohérence d'un ensemble de croyances ne suffit pas par elle-même à justifier que ces croyances soient supposées *correctes*, comme le soutient Davidson, puisqu'un ensemble de croyances fausses peut être cohérent[1]. Mais c'est ici que le PC se trouve étroitement lié au holisme des croyances : en attribuant des croyances à une créature, je ne peux pas faire autrement que supposer que sa structure de croyances est largement similaire à la mienne ; or par définition je suppose mes propres croyances comme vraies ; par conséquent cette structure ne doit pas seulement être supposée cohérente, mais aussi correcte dans son ensemble :

> On n'identifie et ne décrit les croyances qu'au sein d'une trame serrée de croyances. Je puis croire qu'un nuage est en train de passer devant le soleil, mais c'est seulement parce que je crois qu'il y a un soleil, que les nuages sont faits d'eau et de vapeur, que l'eau peut exister sous forme liquide ou gazeuse ; et ainsi de suite, à l'infini. Point n'est besoin d'une liste particulière d'autres croyances pour donner corps à ma croyance qu'un nuage est en train de passer devant le soleil ; ce qu'il me faut en revanche, c'est un ensemble pertinent de croyances qui soient en rapport les unes avec les autres. Si je suppose que vous croyez qu'un nuage est en train de passer devant le soleil, je suppose que vous avez le système de croyances idoine pour renforcer cette croyance-là, et ces croyances que je crois pouvoir vous attribuer doivent, pour venir en renfort, suffisamment ressembler aux miennes pour justifier la description de votre croyance comme une croyance qu'un nuage est en train de passer devant le soleil. Si j'ai raison de vous attribuer la croyance, alors c'est que vous devez avoir une trame de croyances qui ressemble beaucoup à la mienne. Il n'est donc pas étonnant que je ne puisse interpréter correctement vos propos qu'en interprétant de telle sorte que nous puissions nous mettre en grande partie d'accord (1977a : 200, *291*).

L'argument avancé ici par Davidson part de deux prémisses : *a)* c'est une condition de l'attribution d'une croyance que celle-ci soit holistiquement reliée à d'autres, et *b)* c'est une condition de l'attribution d'une croyance que cette croyance et celles auxquelles elle est reliée soient similaires à celles de l'interprète ; et conclut à la nécessité du PC. Il s'agit ici, comme

1. Cf. par exemple Blackburn, 1984 : 239, Engel, 1989 : 116-117.

on l'a déjà dit, du holisme des *croyances*. Mais en vertu du lien intime entre les croyances et les significations, ces deux holismes sont étroitement liés.

Cet argument ne nous permet pourtant que de conclure que nous devons *présumer* que la plupart des croyances de l'interprété sont vraies, en vertu de la nature de la procédure interprétative. Dans d'autres passages, Davidson va plus loin, et soutient que « nous devons tenir comme *acquis* [mes italiques] que *la plupart* des croyances sont vraies » (1975 : 168, *247*) et que l'« erreur massive » est impossible *a priori*. Il va même jusqu'à soutenir qu'il dispose d'un « argument transcendantal » permettant d'établir *a priori* la vérité de la majorité de nos croyances (1973a : 72, *117*). Or c'est une chose de soutenir que les conditions de l'interprétation des croyances justifient une *présomption* de rationalité et de vérité de ces croyances sans laquelle on ne pourrait même pas attribuer de croyances, et c'en est une autre que de dire que ces conditions justifient que les croyances *soient*, de fait, pour la plupart vraies et rationnelles. Davidson (1983) franchit explicitement ce pas en soutenant que l'on peut tirer du PC une réfutation *a priori* du scepticisme. Cette thèse est beaucoup plus forte que la précédente, qui portait sur la nécessité de la charité pour l'attribution de croyances, on doit noter que la méthode d'interprétation proposée par Davidson ne requiert pas en elle-même que tout système de croyance interprété doive être « massivement » correct, ni qu'un interprète ne puisse interpréter un agent que s'il partage, de fait, la plupart de ses croyances (rappelons que c'est la compréhension, et non pas l'accord qui est l'objectif). En d'autres termes, il n'a besoin que du fait que la plupart des croyances *tenues vraies* par l'interprète soient également tenues vraies par l'interprété, et non pas de la thèse selon laquelle ces croyances *sont* vraies[1]. Passer à cette seconde thèse est bien plus audacieux. Dans cette perspective, la théorie de l'interprétation est bien « l'épistémologie au miroir de la signification. » Je n'examinerai pas cet argument avant le chapitre 6.

1. Cf. Bennett, 1985 : 611.

2.4. L'indétermination de l'interprétation radicale

Récapitulons les étapes de la procédure d'interprétation radicale. Cela nous permettra de répondre à la question 4 /, celle de ses conséquences. En premier lieu, l'interprète s'efforce de déterminer le plus grand nombre de phrases que l'interprété tient pour vraies. Appliquant systématiquement le PC, il suppose que le fait que ces phrases soient tenues pour vraies donne une bonne raison de les tenir, de prime abord, pour vraies. Utilisant sa connaissance des objets et des événements de l'environnement, il émet autant d'hypothèses interprétatives, qu'il construit sous la forme de phrases-T, en établissant une corrélation entre les phrases tenues pour vraies par l'interprété et celles qu'il tient lui-même pour vraies. En vertu de PCR, l'interprète essaie également d'imposer le maximum de structure logique de son propre langage au langage de l'interprété. A ce point, l'exigence de donner à la TS pour ce langage la forme d'une théorie-T conduit à une extension importante du PC quinien : alors que ce dernier suppose seulement que la logique des fonctions de vérité peut être appliquée au langage traduit, Davidson requiert d'une théorie-T qu'elle impose aussi et d'abord au langage interprété une structure *quantificationnelle*, parce qu'une théorie-T doit avoir un mécanisme tel que la satisfaction pour les prédicats. Cela suppose en retour que l'on puisse discerner des éléments de structure logique tels que des noms et des prédicats, dans la mesure même où ils sont étroitement liés aux mécanismes de la quantification. Il s'ensuit que l'indétermination de la *forme logique* des expressions, et celle de leur *référence* sont, selon Davidson, beaucoup plus réduites que Quine ne l'admet quand il admet l'« inscrutabilité de la référence ». Je reviendrai sur ce point ci-dessous. La seconde étape discerne les expressions indexicales. La troisième étape traite les autres phrases, celles sur lesquelles il n'y a pas d'accord uniforme, et dont les valeurs de vérité reconnues ne dépendent pas systématiquement de changements dans l'environnement, c'est-à-dire les phrases contenant des termes théoriques (1973 : 136, *203*). Il semble, de prime abord, que cet ordre implique une distinction similaire à celle de Quine entre les phrases « d'observation » et les phrases « théoriques ». Mais Davidson donne beau-

coup moins de poids que Quine à cette distinction, dans la mesure où le fait de discerner d'emblée une structure quantificationnelle entraîne la reconnaissance de prédicats et de noms à un stade que Quine jugeait ultérieur à celui des phrases d'observation (1979 : 230, *332* ; 1983 : 313).

La question essentielle est alors la suivante : la procédure d'interprétation nous permettra-t-elle d'*interpréter* un langage, c'est-à-dire de déterminer effectivement la signification de ses phrases et de le comprendre ? Rappelons-nous l'objection de Foster (§ 1.5) : plusieurs théories-T incompatibles mais extensionnellement équivalentes ne peuvent-elles pas s'appliquer à un même ensemble de phrases ? La réponse dont nous disposons à présent semble être la suivante : les contraintes empiriques de l'interprétation qui viennent d'être énoncées devraient suffire à exclure la possibilité évoquée par Foster. Mais l'excluent-elles réellement ? Non. Davidson admet que, même si les conditions formelles et empiriques sont remplies, un interprète pourra encore forger deux, voire plusieurs, théories-T pour un *L* incompatibles entre elles sans qu'il ait la possibilité de choisir entre l'une ou l'autre de ces théories (1974 : 153, *224* ; 1976 ; 1979 : 224, *325* ; 1991c : 161). Pourquoi est-ce possible, malgré l'addition des conditions empiriques ? Parce, comme le souligne Davidson dans sa réponse à Foster, un interprète pourrait avoir une théorie-T satisfaisant toutes les contraintes, sans savoir que les phrases impliquées par cette théorie *sont* impliquées par cette théorie. Revenant sur les conditions qu'il posait dans (1967), Davidson remarque :

> Mon erreur ne consistait pas, comme Foster semble le suggérer, dans la supposition que (i) n'importe quelle (i) théorie qui donnerait correctement les conditions de vérité serait au service de l'interprétation ; mon erreur consistait à négliger le fait que quelqu'un pourrait connaître une théorie qui soit suffisamment unique sans savoir qu'elle l'est (1976 : 173, *255* ; cf. aussi 1977 : 224, *325*).

En d'autres termes, l'interprète doit a) connaître une théorie-T et les faits qu'elle établit, b) savoir que ces faits ont la forme d'une théorie-T, et c) savoir que cette théorie rencontre les conditions empiriques prescrites, c'est-à-dire qu'elle est « suffisamment unique ». Foster suggère que cette formulation pose un problème pour Davidson, parce qu'elle place, au

nombre des conditions d'une théorie de l'interprétation qui se voudrait *extensionnelle*, une condition *intensionnelle* exprimée par le terme « savoir que », ou par la condition que l'interprète doit savoir que la théorie-T *établit que* ...(tels faits exprimés par la théorie). Il suggère aussi que Davidson pourrait ici utiliser sa propre théorie parataxique du discours indirect (§ 1.5) pour chercher à éliminer cette intensionnalité. Cela suppose en retour que l'analyse en question soit correcte, et cela paraît être une manœuvre bien *ad hoc*[1]. Mais la réponse appropriée est simplement qu'il n'entre pas dans le projet de Davidson d'éliminer *toutes* les notions intensionnelles de sa théorie (comme le montre l'usage de la notion de « tenir-pour vrai »), mais seulement d'éviter celles qui seraient inexpliquées (1976 : 176, *258*).

Mais Davidson a, plus récemment, proposé une autre condition : il exige à présent que les phrases-T produites par l'interprète ne soient pas seulement vraies, mais qu'elles aient aussi la forme de *lois* (1984 : xiv, *11* ; 26, *54*), et même de « lois naturelles » (*ibid.*, xvii, *16*). Cela entraîne l'usage d'une notion intensionnelle, puisque la notion de loi est elle-même intensionnelle (1967c). Mais comment un théorème d'une théorie-T comme « 'La neige est blanche' est vrai ssi la neige est blanche » peut-il être une loi, et une *loi de la nature* ? Outre le caractère mystérieux d'une telle suggestion, elle entre, semble-t-il, directement en conflit avec une autre thèse majeure de Davidson : que le domaine du mental, et par conséquent également celui de la signification, ne peut pas se prêter à la formulation de lois (cf. la section suivante). Il est probable cependant que Davidson pense ici à un concept de « loi » plus faible, d'après lequel les généralisations en forme de phrases-T « supportent » leurs contrefactuels, c'est-à-dire sont des généralisations non strictes, mais suffisamment fortes pour que l'interprète puisse les tenir comme établies. Mais dans ce cas, comme le dit lui-même Davidson (1976 : 175, *258*), pourquoi ne pas formuler ce savoir en disant simplement que l'interprète qui connaît la phrase-T précédente sait que « La neige est blanche » *signifie* que la neige est blanche ? Les « lois » en question ne sont-elles pas alors simplement les *équivalences de signification* ou les synonymies que l'ensemble de la méthode

1. Le Pore et Loewer, 1990, entreprennent d'appliquer systématiquement cette stratégie.

d'interprétation entendait exclure ? Mais il n'y a, là encore, aucune raison de s'alarmer ou de détecter une quelconque circularité[1]. Quand l'interprète sait qu'une théorie-T est « suffisamment unique », il a bien atteint quelque chose que l'on peut appeler une « connaissance des significations » du langage concerné. Le point important est seulement qu'on ait pas *présupposé* cette notion dans l'emploi de la procédure.

Quoi qu'il en soit, même quand l'interprète est parvenu à une théorie « suffisamment » unique, il ne peut pas exclure qu'une autre théorie pourrait se révéler interprétative. En d'autres termes, quel que soit le succès de la procédure, il restera toujours une *indétermination de l'interprétation*, qui peut aussi bien affecter la forme logique que la référence. Davidson nous dit que celle-ci est « l'analogue » de l'indétermination de la traduction chez Quine (1977 : 226, *325*; 1991 : 161). Mais on ne saurait assimiler cette thèse avec IT telle que Quine l'entend, du moins sous sa forme la plus radicale, selon laquelle il n'y a pas de *fact of the matter* quant à la signification. Sous cette forme radicale, IT revient à dire que deux manuels de traduction *incompatibles*, c'est-à-dire divergeant dans l'ontologie qu'ils assignent aux langages, peuvent s'accorder avec les données[2]. Davidson cependant n'accepte IT que sous forme de l'indétermination de la *vérité* et de la *référence* (deux théories-T ou deux schèmes alternatifs d'assignation de la référence peuvent entrer en compétition), mais pas sous la forme de l'indétermination de la signification. Il propose l'analogie suivante :

> Une théorie de la mesure pour la température conduit à assigner des nombres à des objets qui mesurent leur température. Ces théories font peser des contraintes formelles sur les assignations, et doivent être également liées à des phénomènes qualitativement observables. Les nombres assignés ne sont pas uniquement déterminés par les contraintes. Mais le *schème général (pattern)* des assignations est significatif. (Les températures Fahrenheit et Centigrade sont des transformations linéaires l'une de l'autre ; l'assignation de nombres est unique à la transformation linéaire.) D'une manière très voisine, je suggère que ce qui est invariant entre différentes théories acceptables de la vérité c'est la signification. La signification (l'interprétation) d'une phrase est donnée par l'assignation à la phrase d'un lieu sémantique dans le schème général des phrases qui composent le langage. Différentes théories de la vérité peuvent assigner des conditions de vérité différentes à la même phrase (c'est l'analogue sémantique de l'indétermination quinienne de la traduction), alors que les théories sont (manifestement assez) en accord sur le rôle des phrases dans le langage (1977 : 225, *325*).

1. Cf. par exemple Fodor et Le Pore, 1992.

2. Ceci dépend, bien entendu, de la manière dont comprend IT chez Quine. Il existe des versions plus ou moins fortes de la thèse, et bien des variantes, que je n'ai pas analysées ici. Je m'accorde avec Seymour (1992) sur le fait qu'il y a au moins une thèse IT faible, sans doute compatible avec ce que Davidson appelle indétermination de l'interprétation, et une thèse forte, qu'il rejette.

Tout comme on peut mesurer la température en degrés Fahrenheit et en degrés centigrades, on peut mesurer la signification avec différents ensembles de conditions de vérité. Mais tant que la *trame*, ou le schème des assignations est préservé, nous pouvons dire que nous avons saisi la signification comme un invariant. Une analogie comparable peut être faite avec, encore une fois, le cas de la théorie de la décision. Nous pouvons assigner, par exemple, les nombres 0, 1 et 2 comme mesures des valeurs de quelqu'un relativement à l'événement de recevoir 0 F, 5 F et 11 F respectivement. Mais les nombres 2, 4, et 6 conviendraient aussi bien pour établir la même échelle de mesure, puisqu'on remarquerait encore que 6 n'est pas le double de 4. Ce que la mesure établit est une comparaison de différences, et pas de grandeurs absolues (1974a : 146-147, *216*). Si Davidson indique bien qu'il s'agit encore d'une forme d'indétermination, il insiste aussi sur le fait qu'elle est, comparée à celle d'IT selon Quine, parfaitement inoffensive. Il est faux, dans le cas de l'indétermination de l'interprétation, de dire qu'il n'y a pas de *fact of the matter* : « l'invariant est le *fact of the matter* » (1991c : 161). Pour bien voir la différence entre IT selon Quine et la thèse d'indétermination présente, on peut considérer les rôles distincts qu'y joue la thèse du holisme. Chez Quine, comme on l'a vu, le holisme épistémologique est l'une des prémisses conduisant à IT : parce qu'on ne peut vérifier les phrases une à une, leur signification n'est jamais unique. Chez Davidson, le holisme du langage est ce qui permet de garder un invariant en signification. Pour Davidson, le holisme n'est pas la thèse selon laquelle les significations n'existent pas, mais la thèse selon laquelle la nature des faits qui constituent la croyance et la signification n'est pas telle que leur individuation puisse s'effectuer en examinant des croyances et des phrases *isolées*. Alors qu'il revêt chez Quine une forme

négative, le holisme de Davidson prend une forme positive : le fait que pour reconnaître un élément il soit nécessaire de recourir à d'autres est la garantie de la découverte d'une structure, et non pas la menace de la destruction du sens. L'idée qu'il existe une structure commune de la vérité et de la signification nous conduit donc aux antipodes du scepticisme ou du nihilisme de la signification qu'on a parfois prêté à Quine[1].

Davidson considère que les langages humains sont essentiellement intertraduisibles (1973a : 72, *117*). Cette thèse découle non seulement de l'extension considérable qu'il donne au principe de charité, mais aussi de sa conception de l'interprétation : interpréter et communiquer est possible, non seulement parce qu'un fonds de vérités et de croyances est nécessairement commun à l'interprète et à celui ou ceux qu'il interprète, mais aussi parce que l'objectif de l'interprétation est l'accord ou la compréhension. Davidson insiste en outre, comme on l'a vu, sur le fait qu'une théorie de l'interprétation est aussi une théorie empirique, testable. Mais s'ensuit-il cependant qu'on puisse parler de *faits* de signification en un sens comparable à celui dans lequel on parle de faits naturels ? Les formulations initiales de Davidson donnent parfois l'impression qu'une théorie de la signification pourrait être une théorie scientifique, déterminant un univers de faits objectifs. Mais il n'en est rien. Davidson insiste au contraire sur l'irréductibilité de la procédure d'interprétation du langage à une procédure de confirmation empirique du type de celles qu'on peut trouver dans les sciences naturelles. Cette irréductibilité des concepts sémantiques à des faits naturels ou physiques trouve sa source dans une irréductibilité des concepts psychologiques et mentaux aux concepts physiques. Toute la théorie davidsonienne du langage est fondée sur une théorie du mental.

1. Je ne soutiens pas ici cependant que IT chez Quine conduit à ce scepticisme et à ce nihilisme. La distinction entre le holisme quinien et le holisme davidsonien est bien analysée par Heal, 1989, chap. 5. Je soutiendrai au chapitre 7 ci-dessous que le holisme davidsonien s'oppose en tout cas au scepticisme quant à la signification qu'on a tirée de certains arguments wittgensteiniens.

2.5. L'anomie du mental et du sémantique

La question qui nous a occupé jusqu'à présent était celle de savoir comment un interprète peut déterminer les faits qui concernent la signification. La question qui nous occupe maintenant porte sur la *possibilité* de la signification : comment un tel phénomène est-il possible ? Ceci nous permettra d'établir plus précisément le lien entre la théorie de l'interprétation du langage, et la théorie de l'interprétation des croyances. D'un côté la signification est incontestablement un phénomène qui intervient dans un monde naturel et physique, ce qui nous incite à chercher, selon l'expression de David Lewis (1974 : 110), « comment *les faits* [naturels] déterminent les faits [de signification] ». Mais d'un autre côté, les faits de signification semblent échapper à toute réduction de ce type. On se trouve donc confronté à une sorte de dilemme :

> Toute tentative pour comprendre la communication verbale doit la considérer dans son environnement naturel comme faisant partie d'une entreprise plus large. Il semble de prime abord que cela ne puisse pas être difficile, dans la mesure où il n'y a rien de plus dans le langage que des transactions publiques entre des locuteurs et des interprètes, et les aptitudes nécessaires à de telles transactions. Et pourtant cette tâche est hors de notre portée. Car le fait que les phénomènes linguistiques ne soient rien d'autre que des phénomènes comportementaux, biologiques ou physiques décrits dans un vocabulaire exotique où il est question de signification, de référence, de vérité, d'assertion, et ainsi de suite — une simple survenance (*supervenience*) de cette sorte d'un type de fait ou de description par rapport à un autre — ne garantit pas, ou même interdit, la promesse d'une réduction conceptuelle. C'est là qu'est notre problème. Une certaine sorte de réduction semble être requise pour la compréhension, et pourtant toute réduction significative demeure hors de notre portée dans le cas du langage (1980a : 1).

Pourquoi une réduction est-elle « requise » au premier chef ? Parce que les phrases émises par les locuteurs sont des événements physiques, émises par des êtres eux-mêmes physiques, qui ont des croyances et d'autres états mentaux qui sont eux-mêmes identiques à des événements physiques dans leurs cerveaux, parce que ces êtres sont soumis à des lois physi-

ques et sont le produit de l'évolution biologique. Cela suggère que les régularités, règles, voire les « lois » qui semblent gouverner la signification pourraient en dernière instance être réduites à des régularités et à des lois naturelles. Or une telle réduction est, selon Davidson, impossible. Elle est impossible pour au moins trois raisons. En premier lieu, il est impossible de réduire les faits de signification à des faits psychologiques, en raison même de l'interdépendance de la signification et de la croyance : il n'est pas possible d'isoler des faits psychologiques indépendamment des faits de signification, pour déterminer ensuite ceux-ci à partir de ceux-là. En second lieu, et même si on pouvait effectuer le premier type de réduction, il n'est pas possible de réduire les faits psychologiques à des faits biologiques ou physiques plus primitifs, parce que d'une part il n'y a pas de *lois* psychologiques fiables permettant d'expliquer et de prédire le comportement et les actions, et parce que d'autre part il n'y a pas de lois *psychophysiques* permettant de réduire des lois du premier type à des lois du second type. En fait, en vertu de l'interdépendance des croyances et de la signification et de l'impossibilité de lois psychologiques, il n'y a pas non plus de lois « de signification ». Il y a, en ce sens, un « anomisme » (absence de lois) du psychologique et du sémantique. Et pourtant Davidson affirme bien l'identité des événements que sont les énonciations des locuteurs et leurs croyances à des événements physiques et naturels. Il est partisan en ce sens d'un physicalisme et d'un matérialisme *ontologique* (ou d'un *monisme*) d'après lequel il n'y a pas d'autres transactions, quand des locuteurs entrent en communication, que des transactions physiques et naturelles. Mais il rejette l'idée que les *explications* sémantiques puissent être réduites à des explications psychologiques ou physiques, et par conséquent il soutient un antiréductionnisme et un *dualisme* explicatif, d'après lequel il y a une disparité complète entre les explications sémantiques et psychologiques d'une part, et les explications physicalistes de l'autre. Cette position est, dans le domaine sémantique, l'homologue du *monisme anomal* qu'il adopte dans sa métaphysique de l'esprit.

Mon but n'est pas ici de discuter les conceptions de Davidson dans ce domaine, mais seulement d'essayer de tracer — d'une façon nécessairement schématique — les liens les plus significatifs qui existent entre

ces conceptions et sa théorie du langage[1]. La thèse fondamentale de la philosophie de l'action de Davidson est que les actions sont des *événements*, c'est-à-dire des entités individuelles et concrètes, qui sont soumises comme telles aux relations *causales* du monde naturel. Ce qui distingue une action d'un événement naturel est que celle-ci, à la différence de celui-là, peut être décrite comme étant le produit d'une intention, et caractérisée au moyen d'attitudes propositionnelles qui l'expliquent ou en donnent la raison. Les descriptions des actions en termes intentionnels et psychologiques sont intensionnelles alors que les descriptions des événements et des mouvements corporels de l'agent sont extensionnelles. Une autre thèse fondamentale de Davidson est que les explications des actions par les raisons sont des explications *causales* : décrire une action comme accomplie pour telle raison, c'est énoncer sa cause, en mentionnant des états mentaux (désirs, croyances, intentions) qui conduisent l'agent à agir de telle manière. Mais à la différence des théories causales traditionnelles de l'action, cette théorie n'implique pas que les explications causales par les raisons fassent appel à des lois du comportement ou à des lois psychologiques. L'explication causale d'une action est toujours *singulière*, et ne peut reposer sur une science du comportement. Enfin Davidson soutient que l'explication des actions repose sur des principes normatifs de rationalité, comme ceux auxquels fait appel la théorie bayésienne de la décision quand elle suppose que les agents agissent de manière à « maximiser leur utilité espérée », sur la nature desquels on va revenir immédiatement.

La théorie de l'action se rattache nécessairement à une théorie du mental, puisque les actions sont causées par des événements mentaux et expliquées et décrites en termes de ces événements. La thèse du monisme anomal (MA) découle directement des principes qui viennent d'être mentionnés. Un seul et même événement, décrit comme une action, pourra être décrit à la fois sous une description mentale (énonçant sa raison) et sous une description physique (par exemple physiologique), et il entrera dans des

1. Cf. en particulier, pour la théorie de l'action, 1963, 1967a, 1976, 1987a, et, pour la théorie du mental, 1970a, 1973b, 1973c, 1992, ainsi que 1980. J'ai analysé ces théories plus en détail ailleurs, dans Engel, 1986, 1991d, 1992, 1992b, 1993. Pour une excellente confrontation entre les deux théories, cf. Evnine, 1990. L'exposé qui suit négligera bien des distinctions et des précisions indispensables à la cohérence de la thèse, qui ne nous concerne pas directement ici.

relations causales avec d'autres événements, eux-mêmes décrits mentalement ou physiquement. Mais bien que cet événement tombe sous des lois (physiques ou physiologiques) en vertu de sa description physique, il ne tombe pas sous des lois en vertu de sa description psychologique ou mentale. Davidson admet donc 1 / que les événements mentaux sont causalement reliés à des événements physiques, 2 / que les relations causales singulières sont sous-tendues par des lois strictes, mais 3 / qu'il n'y a pas de lois psychophysiques strictes. Selon lui, ces thèses sont compatibles à partir du moment où l'on distingue nettement la *relation* causale entre des événements, qui est indépendante de la manière dont ils sont décrits, de l'*explication* causale de ces mêmes événements, qui *est* dépendante de leurs descriptions. On conclut qu'un événement mental qui cause ou est causé par un événement physique (par 1 /) doit tomber sous une loi physique stricte (par 2 /), et par conséquent que cet événement, qui a une description mentale, a aussi une description physique, et *est* donc un événement physique. Comme on le dit parfois, il y a identité des *événements particuliers* [*tokens*] mentaux et physiques, bien qu'il y ait (en vertu de 3 /) dualité des *propriétés* [*types*] mentales et des propriétés physiques. Les trois thèses, et la conclusion qui en dérive selon Davidson, constituent MA.

La thèse 3 /, l'« anomisme du mental », qui implique une irréductibilité des descriptions psychologiques aux descriptions physiques, implique-t-elle qu'il n'existe pas de relations entre les propriétés mentales et les propriétés physiques ? Non, car bien qu'elle implique qu'il n'y ait pas de relations causales entre ces propriétés, elle n'est pas incompatible avec le fait de soutenir que les propriétés mentales dépendent, de manière systématique, des propriétés physiques. Davidson soutient que les propriétés mentales « surviennent » [*supervene*] sur les propriétés physiques, au sens où il n'y a pas de différence mentale sans différence physique (ou il ne peut y avoir deux événements qui soient identiques sous tous leurs aspects physiques mais différents sous un aspect mental quelconque [1970, 214, 253; 1985, 1992]). Cette survenance [*supervenience*] du mental sur le physique n'entraîne pas la possibilité d'une *réduction* du mental au physique, parce que bien qu'elle implique qu'à toute différence mentale doit correspondre une différence physique, elle n'implique pas que les *mêmes* différences mentales doivent varier systématiquement avec les *mêmes*

différences physiques. C'est en ce sens que le matérialisme de Davidson est « faible » ou « minimal ».

C'est, comme le voit, cette thèse de la survenance qui conduit Davidson à présenter le problème de la signification dans les termes où il les pose dans le texte cité au début de cette section. La fonction de cette thèse est essentiellement ontologique ; elle est destinée à nous garder de toute tentation de conclure du dualisme des *modes d'explication* des événements mentaux et des modes d'explication des événements physiques à un dualisme ontologique pur et simple. Car malgré des déclarations comme celle citée ci-dessus, cette thèse ne joue aucun rôle explicatif ou épistémologique dans la conception davidsonienne de l'explication de l'action ou de l'interprétation du langage. A aucun moment, Davidson n'entreprend d'établir la nature des liens qui pourraient exister entre nos descriptions physiques ou physiologiques du comportement et nos descriptions psychologiques ou sémantiques[1]. Ce qui doit nous intéresser principalement ici, c'est son argument en faveur de l'irréductibilité de ces descriptions les unes aux autres, c'est-à-dire la thèse 3 /.

Dans une large mesure, l'argument en faveur de 3 / dépend de ce qu'il faut entendre par « loi ». Les lois visées ici sont ce que Davidson appelle des « lois strictes ». Une loi stricte est un énoncé conditionnel universel qui a la propriété de ne pouvoir être formulée que dans le vocabulaire d'une seule science (qui est « homonomique ») et de ne pouvoir être précisée ou améliorée en recourant à un autre vocabulaire (auquel cas elle est « hétéronomique »). Par exemple, c'est une loi géologique que les rives supérieures des rivières à méandres subissent l'érosion, sauf si les conditions météorologiques l'empêchent. Cette loi est donc vraie, *ceteris paribus*, moyennant l'exception mentionnée ici dans le vocabulaire de la météorologie. Les lois psychophysiques sont de même nature. Supposons que nous énoncions une telle « loi » supposée : « toute personne déshydratée qui boit ira mieux ». Elle ne vaudra que moyennant des clauses comme : « si elle n'est pas dans un état trop grave », etc. Quant aux lois

1. Ce refus est lié à son rejet des théories « causales » de la référence (cf. chap. 6. ci-dessous). Mais je monterai au chapitre 6 qu'il ne manque pas de poser problème quand il s'agit de parler d'une « connaissance tacite » du langage.

psychologiques, elles ne valent également que moyennant des conditions qui risquent de les annuler ou de les transformer en simples truismes, comme : « Si un individu désire manger une omelette aux fines herbes, alors il le fera en général, si l'opportunité s'en présente et si aucun autre désir ne vient concurrencer le premier » (1973c, 233, *310*)[1]. On ne peut formuler aucune restriction comparable dans le cas des lois physiques. Mais il y a une différence plus profonde encore que la forme même des lois psychophysiques et psychologiques par opposition aux lois physiques. Supposons que l'énoncé précédent ait bien la forme d'une loi. Dans ce cas, pour que l'antécédent soit vrai, il faut que l'individu *désire* manger l'omelette. Mais pour que nous puissions lui attribuer cet état mental, il faut que nous puissions lui en attribuer un ensemble d'autres. Il faut, de plus, que nous supposions que l'individu est rationnel. Ce genre de suppositions n'a pas de sens quand il s'agit d'expliquer un événement physique. En d'autres termes, le holisme du mental et la nécessité de postuler des principes normatifs de rationalité n'ont « pas d'écho » dans la théorie physique (1973c, 231, *308*). Les deux domaines relèvent de principes « constitutifs » qui sont nécessairement distincts (*ibid.*, 221, *299*).

Quels sont les principes normatifs « *a priori* » qui gouvernent le domaine du mental ? Il y a, comme on l'a vu, les principes de rationalité d'action employés en théorie de la décision ou en théorie de la préférence (par exemple la transitivité des préférences (1973c, 1976a), des principes comme le réquisit d'*information totale* du raisonnement inductif d'après lequel un agent croit ce qu'il a les meilleures raisons de croire (1969b), le principe de *continence* (*ibid.*) d'après lequel un agent accomplit l'action qui lui paraît raisonnable sur la base de ses meilleures raisons, et bien entendu le principe de charité. C'est précisément cette nécessité de donner un sens au comportement par rapport à un arrière-plan de raisons, de croyances et d'intentions contraintes par ces principes normatifs qui rendent le domaine mental et le domaine physique « disparates » dans leurs implications (1970a : 222, *300*).

1. Dans 1970 : 216-217, *291*, Davidson lie explicitement cette nécessité d'ajouter des conditions supplémentaires à l'échec des définitions béhavioristes des termes mentaux.

Davidson suggère cependant une autre raison de cette disparité des schèmes explicatifs (1970). Il fait appel à la thèse IT de Quine, et à la version de la thèse de Brentano de l'irréductibilité de l'intentionnel au physique à laquelle souscrit ce dernier (Quine, 1960 : 221) : puisque le domaine du mental est, à la différence du domaine du physique, indéterminé, le premier ne peut être réduit au second. Mais, comme on l'a vu, Davidson est amené à affaiblir IT, et il use lui-même d'une analogie avec la mesure en physique pour exposer le type d'indétermination qui prévaut en théorie de l'interprétation. Il n'est donc pas évident ici qu'IT puisse servir son argument, sauf si l'on comprend cette thèse comme découlant du holisme. C'est donc bien plutôt l'existence de principes normatifs de rationalité qui constitue la source majeure de l'anomisme du mental.

L'anomisme du mental entraîne-t-il qu'il n'y a pas de régularités mentales, pas de lois psychologiques, et par conséquent que la psychologie ne peut pas être une science[1] ? Davidson ne nie pas qu'il y ait des régularités nomiques en psychologie, mais il nie que celles-ci soient strictes. Ce qui distingue la psychologie des autres sciences est le fait qu'elle doit, *à partir du moment où elle se formule en termes d'attitudes propositionnelles*, nécessairement faire appel à des considérations concernant ce qu'il est rationnel pour un agent de faire, employer des considérations quant à la cohérence, à la consistance, et les raisons de ses actions et de ses croyances. Etant donné l'interdépendance de la signification et des attitudes propositionnelles, on peut également parler d'un *anomisme de la signification*. Pas plus qu'il ne peut y avoir de *science* du mental, il ne peut y avoir de science des significations[2]. Cela ne risque-t-il pas d'affaiblir la thèse selon laquelle on peut construire une théorie systématique et empirique de la signification pour une langue naturelle ? Cela ne l'affaiblirait que

1. Cf. les commentaires de R. Peters à 1973c, et la réponse de Davidson (1980, 239-41). Cf également Rosenberg, 1986.

2. Føllesdal (1979) a soutenu au contraire que les hypothèses de rationalité des sciences du comportement pouvaient entrer à titre d'hypothèses générales au sein d'une méthode hypothético-déductive exactement semblable à celle des sciences naturelles. De même Hempel soutient que le principe normatif de rationalité peut prendre la forme d'une hypothèse empirique (cf. Davidson, 1976a). Selon Davidson, ce principe ne peut jamais prendre cette forme.

si l'on supposait, à tort, qu'une telle théorie puisse être une théorie scientifique.

Si MA est cohérent, le « dilemme » mentionné au début de cette section devrait disparaître. Mais il y a néanmoins, comme l'ont remarqué de nombreux critiques, une tension entre l'insistance de Davidson sur le fait que l'explication de l'action est une explication *causale*, qui suppose que les événements mentaux invoqués dans ces explications soient des états *réels* déterminant le comportement, et son insistance sur le fait que ces explications soient essentiellement *interprétatives* et normatives, c'est-à-dire relatives à un observateur. Pour voir comment en quoi consiste cette tension, comparons brièvement la conception de l'interprétation de Davidson avec celles d'auteurs qui défendent une autre conception du mental.

2.6. Réalisme intentionnel et interprétation

Une théorie de l'interprétation des croyances et des contenus d'attitudes propositionnelles vise à répondre à la question : quelles sont les conditions de vérité des phrases rapportant de tels contenus, c'est-à-dire des phrases de la forme « *X* croit que *p* », « *X* désire que *q* », etc. ? La question que nous posons ici est celle de savoir si ces phrases ont des conditions de vérité objectives, c'est-à-dire s'il existe des faits permettant de les rendre vraies ou fausses. Comme on l'a vu, Davidson répond d'abord à cette question en proposant, avec sa théorie du discours indirect, une théorie de la forme logique des comptes rendus d'attitudes : ceux-ci consistent à produire des phrases qui, dans le langage de l'interprète, visent à « dire la même chose » que ce que disent les phrases de l'interprété (§ 1.4). En ce sens le rapporteur des pensées d'autrui, ou l'interprète, doit d'abord avoir accès à ses propres pensées, et aux phrases qu'il tient lui-même pour vraies, avant de pouvoir les assigner à autrui. Comme on l'a vu, le fonctionnement du PC repose sur cette hypothèse. Davidson ne nie pas qu'il y ait un « accès privilégié » de l'interprète à ses pensées, et une asymétrie des pensées « à la première personne » par rapport aux pensées « à la troisième personne » attribuées par un observateur extérieur (1984b, 1987, 1991c). Et parce que l'interprète cherche d'abord à s'établir comme « même

diseur » que celui qu'il interprète, il doit nécessairement chercher à *projeter* le contenu de ses pensées sur celles d'autrui, et par conséquent simuler chez autrui ce qu'il pense chez lui. Mais l'interprétation ne consiste pas en une simple assignation passive de pensées. Elle ne peut se constituer que si l'interprété *répond* aux suggestions de l'interprète. Seulement alors il aura, en les confrontant à ses propres croyances, des critères de similarité entre celles-ci et celles d'autrui. Il n'y a donc que dans le contexte de la *communication* que peut s'établir une interprétation véritable. Ce n'est que dans ce contexte intersubjectif que peut s'établir une *objectivité* de la signification (1982, 480 ; 1991c). Le critère ultime d'attribution des pensées est donc un critère *public*, donné dans la communication, et supposant une communauté d'interprètes.

Malgré cette objectivité, il semble néanmoins que, selon Davidson, la réalité des états mentaux et des significations ne soit jamais à proprement parler « découverte », mais toujours relative aux principes normatifs de l'interprétation elle-même et aux attributions de pensées à la troisième personne qu'il effectue. Il semble ne pas y avoir de *réalité* mentale dont les lois pourraient être étudiées et qu'on pourrait abstraire de ces principes normatifs et holistiques de l'interprétation. En d'autres termes, on peut soupçonner que cette conception conduise à traiter les contenus intentionnels et sémantiques comme seulement des produits de l'interprétation. Il y a, de ce point de vue, une différence importante entre les régularités mentales « hétéronomiques » et celles de la géologie ou de la météorologie : dans ces dernières, nous ne doutons pas qu'il y existe quelque chose de *réel* au sujet de quoi portent ces lois. Or on peut en douter s'agissant de la conception davidsonienne[1].

On peut avoir une tout autre conception des conditions de vérité de nos attributions d'attitudes propositionnelles. Supposons que nous prenions au pied de la lettre la « psychologie populaire » selon laquelle les explications psychologiques des actions sont des explications causales, mais que, contrairement à Davidson, on soutienne — toujours avec le sens commun — que ces explications sont correctes non pas parce qu'elles impo-

1. Les problèmes soulevés dans la présente section m'ont été soumis avec force par Thomas Baldwin. Je les ai abordés ailleurs (Engel, 1986, 1992, 1992).

sent un schème normatif de rationalité sur les agents, mais parce qu'elles font appel à des régularités mentales correspondant à des structures causales *réelles*. Supposons que chaque état mental — une croyance par exemple — s'identifie à une telle structure causale, c'est-à-dire au *rôle* causal qu'il joue, en étant causé par des événements extérieurs au corps de l'agent, en causant d'autres états mentaux, et en ayant des effets comportementaux. Dans ce cas, déterminer le contenu d'un état mental, et l'attribuer à un individu, ce sera identifier un tel rôle causal, et nos conditions d'attributions seront correctes parce qu'elles attribuent de telles structures ou régularités causales aux agents. Cette conception, qui identifie les états mentaux avec leurs rôles causaux, est celle du *fonctionnalisme* contemporain. Selon la version de cette théorie défendue, par exemple, par David Lewis (1970, 1972), on peut construire toutes les généralisations de la psychologie du sens commun portant sur nos croyances, désirs, et autres états mentaux, comme des définitions théoriques implicites dont on pourra donner des définitions explicites[1]. Les états mentaux sont ainsi individualisés par leurs rôles intermédiaires entre des perceptions et des comportements. Selon le fonctionnalisme, aussi bien les états fonctionnels que leurs contenus sont des états *réels* des organismes, même s'ils ne s'identifient pas à des structures physiques. Cette conception paraît bien plus conforme à notre pratique courante d'interprétation dans la psychologie du sens commun : elle suppose, comme cette dernière, que les états mentaux et leurs contenus propositionnels sont des *causes* du comportement, et elle suggère que quand nous interprétons les croyances et désirs d'autrui, nous projetons sur celles-ci une structure similaire à la nôtre propre, conformément au principe d'« humanité » ou de simulation (cf. plus haut). Selon la conception fonctionnaliste, il n'y a pas d'opposition entre le *rôle causal* des états mentaux et leur rationalité. Les structures mêmes qui font qu'un état mental joue un rôle causal, et celles qui font qu'il est rationnel, sont les mêmes structures[2]. Au contraire, chez Davidson, il semble qu'il y ait un divorce entre le *rôle causal* des contenus intentionnels, et leur *ratio-*

1. Selon la procédure de Ramsey d'élimination des termes théoriques. Cf. Jacob, 1992, Engel, 1992, pour une présentation.
2. Cf. Loar, 1981 : chap. 2.

nalité. L'hétérogénéité même des principes normatifs de rationalité interprétative et des structures causales décrites par la physique sur laquelle repose l'anomisme du mental semble interdire de donner un rôle causal aux *contenus propositionnels* des attitudes. Ceux-ci ne sont que des hypothèses postulées par l'interprète, en accord avec des règles de rationalité, pas des structures causales explicatives réelles. Comment, dans ces conditions, Davidson peut-il soutenir, comme il prétend le faire, que les *raisons* d'un agent sont des *causes* de son comportement (1963) ?

Cette objection est renforcée par une objection similaire qui a été adressée par de nombreux auteurs à la métaphysique du monisme anomal[1]. Selon MA, un événement n'est mental qu'en tant qu'il est *décrit* par une description intentionnelle, mais ne rentre dans une relation causale avec un événement physique qu'en tant qu'il est lui-même un événement physique. Comment, dans ces conditions, peut-il avoir un pouvoir causal *en tant que mental* ? Par exemple si Jean désire boire une bière et va vers le frigo, son désir ne cause cette action qu'en tant qu'il est identique à un certain événement physique. Comment peut-on dire alors que le contenu de son désir — *boire une bière* — a causé cette action ? Le contenu intentionnel semble purement inerte ou épiphénoménal. Ou alors il faut se résoudre à dire que le sens du mot « cause » dans la description usuelle en termes psychologiques est distinct du sens de ce mot dans la description physique. Selon MA, l'événement décrit en termes intentionnels ou psychologiques ne tombe pas sous des lois « strictes », alors qu'il tombe sous de telles lois quand il est décrit en termes physiques ou neurophysiologiques. MA ne conduit-il pas ainsi à un véritable dualisme des explications psychologiques (sémantiques) et des explications physiques ? Je ne chercherai pas ici à envisager la réponse que pourrait faire Davidson à cette objection du point de vue de MA[2]. La question qui nous intéresse ici est celle de savoir si, en liant l'individuation des contenus intentionnels aux attributions d'un interprète soumis à des principes idéaux de rationalité, Davidson leur a ôté tout pouvoir explicatif causal, et a défendu,

1. Ces objections sont dues essentiellement à Stoutland, Honderich et Kim. Cf. les articles contenus dans Le Pore, 1986, et pour des références au débat en question, Engel, 1992, 1992b, et Laurier, 1988, 1992.
2. Cf. Davidson, 1992.

comme le soutiennent de nombreux critiques, une conception purement antiréaliste de la nature des contenus mentaux.

Je commencerai par faire une réponse purement dialectique à l'objection « fonctionnaliste » esquissée ci-dessus. Le fonctionnalisme identifie les états mentaux aux rôles fonctionnels qu'ils occupent. Mais en premier lieu, permet-il d'*individualiser* leurs contenus ? Garantit-il, notamment, que deux individus dont les rôles fonctionnels sont identiques auront les mêmes contenus intentionnels ? C'est notoirement douteux, comme le montrent des expériences de pensée célèbres comme celle de la « Terre-Jumelle » de Putnam[1]. Sauf si l'on admet une forme de vérificationnisme, le fait que l'on ne puisse pas individualiser les contenus intentionnels à partir des rôles fonctionnels des états correspondants n'implique pas la fausseté de l'hypothèse fonctionnaliste. Mais on doit admettre que cette dernière n'est pas en meilleure posture, face à l'indétermination des contenus intentionnels, que la position de l'interprète radical. En second lieu, le fonctionnalisme nous garantit-il que les contenus intentionnels ont un pouvoir causal ? Certes, par définition, en vertu de l'hypothèse fonctionnaliste, ils *doivent* avoir un tel pouvoir, puisqu'on les *identifie* à des rôles causaux. Mais cette affirmation a tout d'une pétition de principe, en l'absence d'une explication de la manière dont les rôles fonctionnels peuvent causer un comportement. On peut en fait développer, pour le fonctionnalisme, un argument similaire à celui, énoncé ci-dessus, selon lequel MA serait un épiphénoménisme : un état fonctionnel n'a aucun pouvoir causal, puisque seules ses *réalisations* physiques en ont un ; l'état fonctionnel est en lui-même inerte. Ainsi, pour reprendre un exemple de Lewis, une chaîne d'antivol pour bicyclette a pour fonction d'empêcher le vol de votre bicyclette ; mais elle n'a pas ce pouvoir causal en vertu de cette fonction, mais en vertu des petits disques de métal alignés qui en bloquent l'ouverture ; ce sont eux qui assurent la prévention contre le vol. La menace d'épiphénoménisme n'est donc pas propre à MA, mais à toute conception matérialiste du mental non réductionniste, comme le fonctionnalisme. Enfin, la conception fonctionnaliste de l'interprétation ne

1. Putnam, 1975, 1986. Sur cette objection standard au psychofonctionnalisme de Lewis, cf. Stich, 1983, chapitre 1. Cf. sur Terre Jumelle, ci-dessous § 6.4.

peut réconcilier le caractère causal de nos explications mentalistes usuelles avec leur caractère rationnel que si elle montre que ce dernier dérive, d'une manière ou d'une, de celui-là, par exemple en soutenant que nos organismes ont évolué de manière à incorporer (dans certaines limites) la satisfaction de normes de rationalité dans la structure causale réelle qui détermine leur comportement[1]. Mais même si cette hypothèse est plausible, elle ne montre en rien qu'il pourrait exister un isomorphisme entre les conditions constitutives de rationalité et les structures fonctionnelles et causales des organismes, ni comment cet isomophisme peut être établi. Considérons à cet égard brièvement la conception de la tâche d'interprétation radicale que propose Lewis (1972). Une théorie de l'interprétation radicale pour individu A, selon Lewis, part de la donnée de quatre éléments : 1/ une description physique complète de A comme système physique ; 2/ les attitudes de A telles que nous les exprimons dans notre langage ; 3/ les attitudes de A telles qu'il les exprime dans son langage ; 4/ les significations des phrases du langage de A. Lewis suppose que 1/, le système physique, doit pouvoir déterminer *tout* le reste. Il adopte donc un matérialisme réductionniste, et une hypothèse de survenance forte : « Il ne peut pas y avoir deux individus exactement identiques en 1/, mais différents relativement à 2/, 3/ et 4/. » Lewis admet néanmoins des principes directeurs de l'interprétation qui sont, *prima facie*, semblables à ceux de Davidson : un principe de charité, un principe de rationalité de l'action, et un principe de « générativité » impliquant qu'une structure récursive de conditions de vérité doit être imposée au langage. Notons au passage que cela montre au moins qu'une conception fonctionnaliste-réaliste des contenus mentaux telle que Lewis l'envisage a besoin de principes de rationalité. Mais pour Lewis, ces principes ne sont pas des principes qu'un interprète doit admettre étant donné sa situation particulière ; ce sont des principes qui s'ensuivent de la *définition* même des notions de croyance et de désir. Dans la mesure où ces notions peuvent elles mêmes

1. Je dois cette formulation à T. Baldwin. Elle est implicite dans la défense par Fodor du réalisme intentionnel (Fodor, 1987) : la psychologie populaire, nous dit-il, ne marche pas parce qu'elle est un bon instrument de prédiction du comportement ; elle marche parce que l'espèce a incorporé, au cours de son évolution, cette psychologie à titre de structure *réelle*.

être définies fonctionnellement, charité et rationalité sont elles-mêmes définissables fonctionnellement[1]. Mais Lewis ne nous dit pas pourquoi. Il se contente de le *postuler*.

La conception fonctionnaliste de l'interprétation ne permet donc pas de déterminer quelle relation peut exister entre les principes de l'interprétation des contenus intentionnels et ces contenus eux-mêmes. Mais la question qu'elle soulève est légitime : quelle raison avons-nous de penser qu'il existe des structures intentionnelles réelles derrière les attributions de contenus effectués par un interprète ? Quel est, en particulier, la relation entre le holisme de l'interprétation et le holisme des croyances elles-mêmes ? Davidson a-t-il le droit de dériver le second du premier[2] ? Cette objection est légitime, dans la mesure où Davidson semble adopter dans de nombreux passages un argument purement vérificationniste : l'objet d'une interprétation n'est une pensée, une croyance, ou une action que dans la mesure où cet objet se prête à la méthode d'interprétation. Il semble parfois défendre ce que l'on appelle « interprétationnisme » ou « ascriptivisme », la thèse selon laquelle les contenus intentionnels ne sont *rien d'autre* que des entités postulées utiles à l'interprétation, qui est une forme d'antiréalisme intentionnel[3]. Ce genre d'objections, avancées du point de vue d'un *réalisme intentionnel*, appellent deux sortes de réponses. La première est qu'un réalisme intentionnel radical paraît difficilement tenable. La seconde est que Davidson n'est pas un interprétationniste au sens incriminé.

Selon la conception réaliste radicale des états intentionnels défendue, en particulier, par Fodor (1987, 1990), les croyances, les désirs et les autres états mentaux intentionnels des individus sont des entités totalement objec-

1. Fodor et Le Pore, 1992, chapitre 4, en concluent, à juste titre, que Lewis doit soutenir une conception holistique des croyances, indépendante, contrairement à celle de Davidson, des conditions d'attribution ou d'interprétation. Pour une conception semblable à celle de Lewis, cf. Loar, 1981. Et pour une réponse de type davidsonien à Lewis et Loar, cf. Mc Dowell, 1985.

2. C'est en particulier l'une des objections de Fodor et Le Pore, 1992.

3. Cette thèse instrumentaliste est généralement attribuée aussi à Dennett (Dennett 1987). Cf. Engel 1992, chapitre 4. Cf. par exemple Davidson, 1980a, 6 ; 1985b, 92 : « Ces maximes d'interprétation ne sont pas simplement des conseils utiles ou amicaux : elles sont plutôt destinées à externaliser et formuler (sans doute très grossièrement) des aspects essentiels des concepts communs de penser, d'affect, ou de raisonnement. *Ce qui ne pourrait pas être le produit de ces méthodes ne serait pas de la pensée, du discours, ou de l'action* » (mes italiques).

tives et isolables. Qu'elles soient *objectives* veut dire qu'elles existent indépendamment des structures rationnelles d'attributions de contenus que doit, selon Davidson, présupposer tout interprète pour fixer la nature et les contenus mêmes des états intentionnels, et que leur pouvoir explicatif et prédictif pour le comportement est indépendant de ces présuppositions de rationalité. En d'autres termes, elles entrent dans des *lois* intentionnelles qui peuvent être découvertes indépendamment de toute hypothèse de rationalité interprétative. Qu'elles soient *isolables* veut dire qu'elles sont des états discrets, qui peuvent être fixés indépendamment les uns des autres. En d'autres termes, une croyance est pour Fodor un état *atomique*, dont l'existence et la détermination ne dépendent pas de l'existence ni de la détermination d'autres croyances et états de la créature qui l'entretient. C'est de plus un état *interne* de l'esprit du sujet, impliquant une relation à une phrase de son « langage de la pensée » ou « Mentalais » inscrite quelque part dans son cerveau, et dotée d'une syntaxe et d'une sémantique indépendantes. Ces trois hypothèses s'opposent terme à terme à celles sur lesquelles s'appuie la conception davidsonnienne de l'esprit et du langage. Je n'examinerai pas ici la troisième hypothèse (mais cf. § 6.4) : elle implique ce que l'on appelle, dans la philosophie contemporaine, une forme d'*internalisme* selon lequel le contenu des états mentaux est déterminé, au moins en partie, par les propriétés internes de la syntaxe des symboles mentaux. Les deux premières hypothèses s'opposent explicitement au holisme de Davidson, selon lequel, comme on l'a vu, on ne peut attribuer une croyance isolée sans en attribuer une quantité d'autres, rationnellement reliées à la première. Fodor soutient non seulement une forme d'*atomisme* selon lequel une créature pourrait très bien avoir *une seule* croyance, mais nie également que le contenu des croyances dépende de leurs liens rationnels à d'autres contenus. La plausibilité de sa thèse repose dans une large mesure sur sa conception positive de la nature des contenus mentaux, que je n'examinerai pas ici[1]. Si l'on se concentre par conséquent sur les deux premières hypothèses, à quoi revient l'atomisme de Fodor ? Il revient à l'idée que nous pourrions, par exemple, découvrir que quelqu'un croit que *p*, et qu'il croit que *si p alors q*, et que nous

1. Cf. Fodor, 1987, 1990.

pourrions déterminer, *indépendamment* de l'attribution de ces croyances, s'il croit ou pas que *q*, et que notre attribution de la croyance que *non q* ne menacerait pas notre attribution des croyances *p* et *si p alors q*. Selon Davidson, c'est impossible : nous ne pouvons pas attribuer à un agent de telles croyances sans présupposer qu'il admet un principe logique tel que le *modus ponens*. Si nous devions découvrir que l'agent ne croit pas que *p* et que *si p alors q*, cela devrait être dans un contexte où nous pouvons aussi attribuer à l'agent la croyance que *non q*[1]. C'est l'un des sens dans lesquels l'attribution de ces croyances est « holistique » : l'attribution de croyances isolées suppose que l'agent soit capable d'effectuer des inférences logiques élémentaires, dans lesquelles entrent d'*autres* croyances, inférentiellement liées aux premières en vertu de principes logiques normatifs. Notons que c'est ici le principe de charité en tant que principe de cohérence (PCR) et non en tant que principe de véridicité (PCV) qui intervient[2]. Selon Fodor au contraire, on devrait pouvoir avoir des confirmations de l'attribution de ces croyances indépendantes de tout principe normatif de rationalité logique : en identifiant dans la tête des agents une « boîte à croyances », et en identifiant dans cette boîte les phrases particulières du Mentalais. Cela suppose que l'on puisse identifier la syntaxe des états mentaux et leur sémantique indépendamment de leur interprétation dans le cadre d'un langage public, et, compte tenu des conditions de l'interprétation radicale, indépendamment des normes de rationalité que celle-ci incorpore. C'est certes faire une pétition de principe contre Fodor que de supposer que l'on ne puisse pas fixer les contenus d'états intentionnels en dehors des conditions d'une interprétation radicale. Mais c'en est une aussi de la part de Fodor que de supposer le contraire. La seconde hypothèse essentielle de Fodor est qu'il peut y avoir des confirmations indépendantes de l'attribution de croyances individuelles

1. Cf. Dennett, 1993, p. 217, qui répond à Fodor et Le Pore, 1992, chap. 5.

2. Comme le fait remarquer Laurier (1994), le principe de charité sous sa forme PCV n'est pas nécessairement une prémisse conduisant au holisme, même si la supposition que la *plupart* des croyances d'un agent sont vraies implique qu'un agent qui aurait une seule croyance (un esprit « ponctué », comme le dit Fodor) ne pourrait pas être interprété. Mais sous sa forme PCR, invoquée ici, il semble bien impliquer le holisme, puisque toute attribution de croyances rationnelles suppose que l'agent soit capable d'effectuer diverses inférences rationnelles, et par conséquent qu'il effectue des transitions entre *diverses* croyances. Je reviens sur ces points au § 6.5.

si ces attributions sont étayées par des lois intentionnelles elles-mêmes bien confirmées. Elle revient donc à nier explicitement l'une des prémisses de MA, l'anomie du mental. La question de savoir si cette prémisse est correcte est évidemment très débattue[1]. Comme on l'a vu, Davidson ne nie pas qu'il existe un certain nombre de régularités, formulables en termes contrefactuels et intentionnels, qui méritent d'être appelées des « lois », et qui ont, comme Fodor l'admet lui-même, *ceteris paribus*. Le problème est de savoir si ces régularités ont, comme le soutient Fodor, un caractère *empirique*, et si elles sont énonçables indépendamment d'hypothèses portant sur la rationalité des agents. Le problème est qu'il est très difficile d'énoncer ce genre de lois sans faire appel d'une manière ou d'une autre à un élément normatif. Supposons par exemple que l'on dise que c'est une « loi » que

> *Ceteris paribus* si X perçoit qu'il y a un cube en face de lui, alors X croira qu'il y a un cube en face de lui.

En quoi cette « loi » est-elle distincte de

> *Ceteris paribus* si X *devait croire*, sur la base de sa perception, dans des conditions normales, qu'il y a un cube en face de lui, alors X croirait qu'il y a un cube en face de lui.

Il n'est pas facile de le dire, parce que nos attributions de croyance sont étroitement dépendantes de notre évaluation de la situation comme « normale », des croyances de l'agent comme liées à d'autres croyances rationnelles qu'il devrait avoir, en d'autres termes à des conditions normatives d'attribution[2]. Ici encore, le holisme défendu par Davidson repose sur la thèse du caractère normatif de toute attribution d'intentionalité et de signification. Il n'y a certes pas de lien direct entre cette thèse et la thèse holiste selon laquelle on ne peut pas attribuer à un sujet une croyance sans lui en attribuer un certain nombre d'autres. Mais si l'on

1. Cf. nombre des articles dans Le Pore, 1985.

2. C'est en ce sens, me semble-t-il, que Peacocke (1986) parle des « engagement canoniques » que doit avoir tout sujet qui entretient certains contenus. Une partie importante de la thèse de Peacocke porte sur l'explicitation de ces conditions normatives d'acceptation pour diverses variétés de contenus de pensée. Cf. ci-dessous, § 5.5.

admet que a) la normativité des concepts impliqués dans l'interprétation radicale est l'une des raisons principales pour lesquelles b) il n'y a pas de lois intentionnelles strictes et que c) toute attribution d'une croyance présuppose que l'agent soit capable de faire certaines inférences rationnelles, alors on a de bonnes raisons de soutenir la thèse holiste sous la forme énoncée ci-dessus, à savoir d) qu'*on ne peut attribuer une croyance isolée sans en attribuer d'autres, rationnellement reliées à la première*. C'est parce que l'application des normes de rationalité implique nécessairement qu'un agent ait un ensemble de croyances inférentiellement reliées que toute attribution d'intentionalité implique l'attribution d'un ensemble de croyances. Davidson admet lui-même[1] que ces considérations ne constituent pas un *argument* en faveur de la forme de holisme qu'il défend, parce que les raisons a), b) et c) ne sont pas indépendantes les unes des autres. Il reste aussi beaucoup à préciser quant à la notion de normativité invoquée ici, et quant au holisme (j'y reviens au § 6.5)[2]. Tout au plus ces considérations rendent-elles plausible la conception davidsonnienne de l'interprétation, et peu plausible une forme de réalisme intentionnel extrême comme celui de Fodor.

Nous n'avons cependant pas encore répondu à la question qui motivait l'objection de l'« interprétationnisme » attribué à Davidson. En quoi le fait que l'interprétation obéisse a de telles contraintes implique-t-il quoi que ce soit quant à la *nature* même des croyances et des significations attribuées selon ces règles ? Ces règles ne pourraient-elles pas s'appliquer quand bien même nous n'aurions pas de croyances, de désirs, ni d'états intentionnels ? Davidson ne dit-il pas lui-même que des états ne sont intentionnels qu'*en tant que décrits* ou interprétés d'une certaine façon ? Sa position n'est-elle pas alors très proche des formes d'antiréalisme quant aux états intentionnels que l'on décrit comme *instrumentalisme* (les états mentaux n'existent qu'en tant que postulats heuristiques de l'interprétation permettant de prédire le comportement) ou comme *éliminativisme* (les états mentaux n'existent pas)[1] ? La réponse est que Davidson n'est un antiréaliste intentionnel en aucun de ces sens, bien qu'il soit légitime de considérer que sa position est une forme d'antiréalisme, si on la compare aux diverses thèses réalistes intentionnelles de la philosophie contemporaine. Je reviendrai, au chapitre 6, sur ce point, mais tout dépend évidemment de ce que l'on doit ici appeler « réalisme » quant au mental. Il y a deux variantes extrêmes de cette thèse. La première est celle que nous venons de mentionner, dont Fodor est l'un des représentants : un réalisme intentionnel naturaliste, selon lequel les états mentaux s'identifient à des structures objectives du cerveau, dont les contenus sont déterminés par des relations causales avec l'environnement[2]. La seconde est la forme de cartésianisme défendue par Searle (1983, 1992), selon laquelle la réalité des états mentaux est directement accessible, à la première personne, par les agents eux-mêmes. Selon Searle, toute détermination des contenus mentaux par une procédure d'interprétation doit reposer sur le point de vue « à la troisième personne » d'un interprète qui attribue ces contenus de l'extérieur, et par conséquent réduit à de simples postulats de l'interprète. Il est clair que Davidson n'est un réaliste intentionnel ni au premier sens, ni au second. Le caractère spécifique de son antiréalisme, ou de son réalisme, tient à la question de savoir en quoi la *nature* et *l'objectivité* des croyances et des significations est supposée *dépendre* de leur interprétation. Davidson est clair sur ce point : une croyance ne serait pas une croyance si elle n'était pas interprétée selon la procédure qu'il envisage. Mais cela n'implique pas que croyances et significations n'existent qu'*en tant qu'*elles sont interprétées, comme le voudrait la thèse « interprétationniste » telle qu'on l'a envisagée. C'est parce qu'il y a quelque chose à interpréter que l'interprétation est possible. Comme on l'a vu (§ 2.4),

1. Par exemple oralement au cours d'une table ronde consacrée à son œuvre en 1992 dont les actes sont réunis dans Engel, 1994.

2. On doit noter en particulier que ce n'est pas en soi le caractère « normatif » des attributions d'intentionalité qui justifie l'anomisme du mental, puisque selon Davidson (1970) le domaine du physqiue est lui aussi régi par des « normes ». Ce qui justifie cet anomisme est que les normes qui régissent le mental ou l'intentionnel sont, selon Davidson, foncièrement distinctes des normes qui peuvent régir le physique.

1. La première doctrine est couramment attribuée à Dennett, qui la répudie au nom de considérations proches de celles développées ci-dessous (Dennett, 1987), la seconde à P.M. et P.S. Churchland. Cf. Engel, 1992, chapitres 2 et 4. Cette objection fait surface à de nombreuses reprises chez les critiques de Davidson.

2. La version du naturalisme de Fodor n'est pas la seule possible. D'autres versions sont celles de Dretske (1988) qui définit les contenus mentaux en termes de certaines structures causales informationnelles, et celle de Millikan (1984) qui les définit en termes de structures fonctionnelles-téléologiques. Cf. Engel, 1992, chapitre 5.

Davidson soutient que l'interprétation du langage et du comportement dégage des *trames* objectives et des invariants que l'on peut tenir comme des *faits* objectifs indépendants des moyens que nous avons de les *mesurer*. Le caractère indéterminé de nos interprétations des pensées, des significations et des actions n'implique pas qu'il n'y ait rien à interpréter. En ce sens, Davidson n'est pas un antiréaliste quant à l'intentionnalité. Il ne s'oppose pas à un réalisme intentionnel, mais aux versions fortes de ce réalisme, selon lequel les contenus intentionnels seraient objectivement déterminés indépendamment de nos procédures interprétatives.

Une « trame », une « structure », ou un « invariant » sont des entités plus évanescentes que des représentations propositionnelles ou des symboles dans le cerveau. Mais tant que l'existence de ces dernières n'a pas été établie, on voit mal en quoi on peut se prévaloir d'un réalisme plus fort. Néanmoins l'existence de trames ou de structures communes dans l'interprétation des individus n'est pas une preuve de leur objectivité. Pourquoi une structure sur laquelle s'accordent les interprètes devrait-elle être objective ? Et même si la communication présuppose, comme le soutient Davidson, une norme objective de vérité, pourquoi celle-ci serait-elle l'unique norme possible ? A ceci Davidson répond que la source de l'objectivité est une « triangulation » entre deux personnes et une source de stimuli. Les deux agents notent leurs réactions et les relient entre elles (1982, 1991). Le critère de l'objectivité est donc intersubjectif : il n'est possible que s'il y a communication. Nous ne pouvons pas, ajoute Davidson, chercher un critère plus fondamental que celui-là, parce que la correction du critère lui-même dépend du processus d'interprétation : « Une communauté d'esprits est la base de toute connaissance ; elle fournit la mesure de toutes choses » (1991c : 164). Cette communauté est le fondement de la normativité des concepts mentaux, et la raison ultime de la différence entre ces concepts et ceux qui s'appliquent à l'univers physique. Il est correct de dire, en ce sens, que pour Davidson la réalité et la nature des contenus mentaux « dépendent » de leur interprétation, mais il est faux d'en conclure que cette réalité se *réduit* à celle d'hypothèses ou de postulats d'un interprète. Par définition, il n'y a pas d'interprétation sans communication, et pas de communication sans partage d'un univers objectif extérieur aux interprètes. La réponse générale de Davidson

à l'objection selon laquelle il défendrait un antiréalisme « interprétationniste » repose donc sur l'idée qu'il y a convergence dans l'interprétation, et que cette convergence est la seule garantie de l'objectivité de ce qui est interprété. Nous verrons plus précisément, au chapitre 6, quelle est la nature de ce « réalisme ».

2.7. Une théorie « généralisée » du langage et de l'action

L'objectivité d'une théorie de l'interprétation dépend en partie de la manière dont on peut *confirmer* une interprétation du langage, des pensées et des actions d'un agent. La réponse à cette question ne dépend pas d'un argument « transcendantal » comme celui que nous avons évoqué ci-dessus. Davidson ne renonce pas à sa thèse initiale de 1967, selon laquelle une théorie de la signification est une théorie empirique *testable*, qu'on puisse rattacher, de manière spécifiable, à l'*usage* des phrases par des locuteurs. Comme nous l'avons vu (§ 2.2), il conçoit ce problème comme étant fondamentalement le même que celui d'une théorie de la mesure des degrés de croyances et de désirs dans la théorie classique de la décision. La solution qu'il cherche à apporter à ce problème est également inspirée de celle-ci et prend la forme d'une « théorie unifiée du langage et de l'action » (1980, 1985, 1990). Le reproche principal que fait Davidson à la théorie classique de la décision est qu'elle ne tient pas compte, dans son évaluation des désirs et des croyances, des significations que les locuteurs accordent aux réponses qu'ils donnent à l'expérimentateur (cf. 1976a) : elle présuppose en fait que l'on puisse individualiser les propositions qui sont les contenus des croyances et des désirs des agents *indépendamment* de l'évaluation des degrés de ces croyances et désirs. Il lui manque donc une théorie de l'interprétation du discours. Davidson propose donc d'intégrer dans une structure commune ces trois éléments, en vue de chercher à déterminer simultanément des degrés de croyance, des degrés de désir, et des significations, sans présupposer ni les uns ni les autres. La stratégie est donc le même que celle d'une théorie de l'interprétation. Le problème sera, comme précédemment, de choisir une base empirique suffisante. Davidson propose de prendre comme donnée ini-

tiale l'attitude de *préférer-vraie* une phrase plutôt qu'une autre. Cette attitude est fonction de la signification que l'agent assigne aux phrases, de la probabilité qu'il attache aux états de choses dont la vérité des phrases dépend. Comme la notion initiale de *tenir-pour-vrai*, on doit pouvoir savoir qu'un agent a cette atttitude sans savoir ce que ses phrases signifient, sans savoir non plus quels sont ses degrés d'utilité et de croyance. Mais Davidson modifie de manière importante la théorie de Ramsey. En premier lieu, la théorie de Ramsey fait un usage essentiel des paris, comme bases des préférences. Or comment peut-on savoir qu'un agent tient une phrase comme présentant un pari tant que nous ne savons pas comment interpréter son langage ? Davidson propose ici d'adopter plutôt la version de la théorie de la décision de Jeffrey, qui n'utilise pas directement la notion de pari, mais qui traite les objets de préférence comme des *propositions*, et extrait les probabilités subjectives et les utilités des propositions tenues pour vraies[1]. Il en résulte une diminution du degré de précision de la théorie ; mais celle-ci reflète, selon Davidson, l'indétermination inévitable de l'interprétation. Mais les propositions sont tenues comme les significations des phrases : il faut donc reformuler les préférences comme des préférences par rapport à des phrases. La seconde étape consistera à exprimer les connecteurs vérifonctionnels en termes de préférer-vrai[2]. Une fois les connecteurs identifiés, on peut fixer les degrés de croyances et de désirs. Davidson soutient donc qu'on peut déterminer la structure vérifonctionnelle des phrases d'un agent à partir de ses préférences, à partir de là, selon la méthode usuelle d'interprétation, leur contenu propositionnel. La procédure est ingénieuse, puisqu'elle incorpore à présent non seulement les croyances et les significations, mais aussi les désirs et les préférences de l'agent. Comme les croyances, les préférences sont soumises à des normes de rationalité (comme la transitivité : si un agent préfère A à B et B à C, alors il doit préférer A à C) et en ce sens la procédure

1. Cf. Jeffrey, 1965 ; Davidson, 1980 : 10.

2. Davidson montre qu'on peut partir d'un opérateur O dont l'application à deux phrases renverse l'ordre de préférence : si *s* est préféré à *s'*, alors Os' est préféré à *s*. On détermine ensuite l'opération sur deux phrases *s* et *s'* telle que *s* est préférée vraie à *s'* si et seulement si le résultat de l'opération sur *s'* et *s'* est préféré au résultat de l'opération sur *s* et *s*. Cette opération est le « ni...ni » de la logique des fonctions de vérité. Les autres connecteurs peuvent être déterminés à partir de là.

paraît être une simple extension de la méthode initiale. Mais en un autre sens, elle accroît considérablement la base empirique de la méthode d'interprétation, puisqu'elle suppose non seulement que les agents tiennent des phrases comme vraies, mais qu'ils les tiennent pour vraies à un certain degré, relatif au degré de leurs désirs. Davidson ne montre pas comment nous pouvons effectuer ces assignations de manière détaillée, par plus qu'il ne nous indique comment nous pouvons repérer une attitude de préférence vis-à-vis de phrases[1]. Mais l'intérêt de cette reformulation de la procédure d'interprétation en termes directement « bayésiens » est avant tout théorique. Elle est destinée à montrer qu'il existe une interdépendance des attitudes propositionnelles d'un agent telle qu'on ne peut espérer en déterminer une indépendamment des autres. Comme une théorie de la mesure de grandeurs physiques, une théorie de l'interprétation repose sur la postulation d'une certaine *structure* dont les concepts de vérité et de préférence font nécessairement partie.

On prête en général peu d'attention à ces considérations davidsoniennes sur le rôle de la théorie de la décision dans l'interprétation. Mais elles occupent une place essentielle pour qui veut comprendre son holisme et la manière dont sa théorie du langage et sa théorie de l'action rationnelle entrent en contact. La théorie bayésienne de la décision joue en fait pour lui le même rôle que celui que joue la théorie de la vérité de Tarski : celle d'un instrument de mesure du mental. De même que celle-ci détermine la structure des contenus sémantiques, celle-là détermine la trame des attitudes propositionnelles reliées.

1. Dans 1990, note 67, p. 325, il admet que cette méthode est « artificielle ».

3

Interprétation et communication

> Thorbjoern Berntsen, ministre de l'Environnement, a confirmé hier ses propos. Lors d'une réunion électorale, lundi, il avait accusé les Britanniques de provoquer par leur pollution des pluies acides en Norvège, précisant que John Gummer, son homologue britannique, était « le plus beau sac de merde qu'il ait jamais rencontré ». Hier Berntsen a avoué qu'il « aurait pu employer d'autres termes », mais qu'en général il préférait « se faire comprendre clairement ».
>
> *Libération*, 18 août 1993

3.1. Signification, usage et contexte

L'une des objections les plus fréquentes que l'on adresse à une conception vériconditionnelle de la signification comme celle de Davidson est qu'en faisant de la vérité la dimension principale d'évaluation du sens, elle repose au mieux sur une idéalisation, en abstrayant les conditions de vérité des phrases de leur contexte et — dans l'un des sens de cette notion — de leur usage effectif par les locuteurs. Or d'un côté la plupart, sinon la majeure partie des phrases d'une langue naturelle, ne sont vraies que relativement aux contextes dans lesquelles elles sont énoncées, ne serait-ce que parce qu'elles contiennent des expressions indexicales. Et de l'autre l'énonciation de phrases vraies n'est qu'une des multiples choses que des locuteurs peuvent effectuer grâce au langage, puisqu'ils peuvent également donner des ordres, poser des questions, exprimer des souhaits ou adresser des requêtes. Ces deux dimensions, l'indexicalité et la force illocutoire, sont couramment celles dont traite une « pragmatique », par opposition à une syntaxe et à une sémantique, et l'objection revient donc à dire qu'une théorie de la signification au sens davidsonien

ne s'occupe que des secondes à l'exclusion de la première. L'objection est injustifiée. Comme on l'a vu, Davidson insiste sur le fait qu'en raison de l'indexicalité, une théorie de la vérité appliquée à une langue naturelle devra relativiser les phrases-T aux locuteurs, aux lieux et aux temps. Et sa conception de l'interprétation radicale est précisément destinée à rendre compte de l'interprétation des phrases dans des occasions d'usage particulières. Le problème n'est donc pas que cette conception passerait sous silence ces conditions pragmatiques de la signification, mais il est celui de savoir si elle permet d'en rendre compte de façon satisfaisante. Si l'on accepte, au moins provisoirement, la distinction traditionnelle entre une sémantique et une pragmatique, la difficulté rencontrée par la stratégie de Davidson au sujet de la première — comment rendre compte, à partir d'un cadre théorique volontairement « austère » ou pauvre, d'un donné « riche » ? — se pose encore plus nettement pour la seconde, dans la mesure même où les théories pragmatiques font appel à des notions comme celles d'intention, de convention, de contexte, de force illocutionnaire ou d'acte de langage. Comment Davidson peut-il continuer à appliquer le même minimalisme conceptuel face à des phénomènes linguistiques qui paraissent défier ce genre d'entreprise ? Sa stratégie doit être double. En premier lieu, il doit montrer qu'au plan *sémantique* les expressions indexicales, les phrases à l'impératif, ou les performatifs, qui semblent résister à une incorporation dans le cadre d'une théorie de la vérité, sont bien susceptibles d'entrer dans ce cadre. Cela apparaît de prime abord difficile. Comme on l'a vu, pour les phrases indexicales, Davidson propose que les phrases-T aient la forme suivante :

(1) « Ce livre a été volé » est vrai en tant que (potentiellement) prononcé par *p* en *t* ssi le livre montré par *p* en *t a été volé avant t* (1967 : 34, *66*).

qui, de son aveu même, peut difficilement passer pour une analyse[1]. L'histoire récente de la sémantique des démonstratifs semble montrer que la simple relativisation de la vérité des phrases les contenant à divers paramètres est insuffisante pour rendre compte de leurs conditions de vérité,

1. Le problème, comme le dit Davidson, est qu'avec des phrases-T comme (1) contenant des démonstratifs, « le côté droit du biconditionnel ne traduit jamais la phrase dont il donne les conditions de vérité » (1976 : 175, *258*).

et que si elles se prêtent à un traitement vériconditionnel, c'est bien plus facilement dans le cadre d'une sémantique intensionnelle, ou dans celui d'une théorie sémantique « riche », et non pas dans le cadre d'une théorie-T[1]. Le problème semble encore plus épineux dans le cas des phrases impératives ou performatives, car

(2) « Viens au cinéma » est vrai ssi viens au cinéma
(3) « Je m'excuse » est vrai ssi je m'excuse

ne peuvent pas servir de base à une sémantique adéquate, tout simplement parce les phrases prises pour cibles ne semblent se laisser évaluer en termes de leurs seules conditions de vérité[2]. En second lieu, Davidson doit montrer que les concepts additionnels dont on a besoin pour faire d'une théorie de la vérité une théorie de la signification sont bien ceux de sa théorie de l'interprétation radicale, et seulement ceux-ci. En d'autres termes, pour employer une distinction de Dummett que lui-même emprunte à Frege, une théorie de la signification peut être décomposée en une théorie du *sens* et une théorie de la *force*[3]. La première analyse les contenus sémantiques des phrases, et la seconde identifie les actes linguistiques (assertions, ordres, questions, etc.) effectués avec ces contenus. Comme on l'a vu, une théorie-T doit pouvoir servir, dans les conditions appropriées, de théorie du sens. La question est donc de savoir quels concepts doit utiliser une théorie de la force.

Dans ce qui suit, je laisserai de côté le problème des indexicaux, pour lequel les suggestions de Davidson sont manifestement insuffisantes, et je ne considérerai que le cas des modes, qu'il développe plus abondamment. Mais avant d'aborder le problème sémantique, considérons plutôt

1. Cf. Lewis, 1972, Montague, 1974, et surtout Kaplan, 1977, pour la sémantique intensionnelle des démonstratifs. La plupart des analyses récentes, comme celles de Evans (1981) ou de Perry (1979) réintroduisent des notions comme celle de sens frégéen ou des notions apparentées, et usent d'un appareillage conceptuel bien plus complexe que celui requis par Davidson, ou par des tentatives inspirées par son programme (Weinstein, 1974, Burge, 1972). On notera que les mêmes difficultés se poseraient pour l'analyse des temps verbaux. J'ai touché quelques-uns de ces problèmes dans Engel, 1991b.

2. (3) Paraît être évaluable, mais laisse précisément de côté la force, en sorte qu'il est douteux que les conditions de vérité de la phrase de gauche, si elle en a, soient celles qu'énonce la phrase de droite.

3. Dummett, 1973 : 416 ; 1976 : 74 ; cf. aussi Mc Dowell, 1976 : 44, Davies, 1981 : 7-9.

la manière dont on pourrait formuler une théorie de la force, et essayons de le faire à partir d'un cadre conceptuel qui ne soit *pas*, de prime abord, celui de Davidson, celui que proposent des auteurs comme Grice.

3.2. Signification et intentions de communication

Selon Grice (1957, 1969), il y a deux composantes de la signification :

(a) la signification *du locuteur* (ou *l*- signification) : le fait qu'un locuteur signifie que *p* au moyen d'une phrase *s* est définissable comme un certain type de comportement effectué avec l'intention d'activer la croyance que *p* chez un auditeur (de communiquer que *p*) ;
(b) la signification *linguistique* ou *de l'expression* (*e*-signification) : la signification d'une phrase *s* est définissable comme certaines corrélations entre des marques et des sons et des types d'actes de signification du locuteur[1].

Cette analyse est supposée nous fournir une définition, ou une réduction du concept de signification linguistique à celui de signification du locuteur : une phrase *s* signifie, dans des circonstances données, que *p* si et seulement si le locuteur qui l'énonce a l'intention d'activer chez un auditeur la croyance que *p*. Autrement dit, il s'agit d'analyser la notion de signification en termes d'attitudes propositionnelles des locuteurs et des auditeurs, et par conséquent de fournir une certaine forme de réduction des concepts sémantiques à des concepts psychologiques.

Les deux composantes — signification linguistique et signification du locuteur — sont, chez Grice, réunies au moyen de la notion de *convention* : on suppose qu'il y a, associées à des actes de signification du locuteur, certaines conventions par lesquelles les locuteurs effectuent ces actes avec certaines intentions. La notion de signification du locuteur est elle-même définie ainsi :

(1) un locuteur *S* a l'intention que *s* produira chez un auditeur une croyance que *p*
(2) pour un trait particulier *T* de *s*, S a l'intention que *A* reconnaisse l'intention première (celle de (1)) en partie en reconnaissant que *s* a le trait *T* ;

1. Je suis ici les formulations de Schiffer, 1987 : 242 ; cf. également, Bennett, 1976.

(3) *S* a l'intention que la reconnaissance de l'intention première de *S* soit une partie de la raison qu'a *A* de croire que *p*
(4) *S* n'a pas l'intention que *A* soit trompé par les intentions de *S*.

A cette définition on doit ajouter des notions d'actes illocutionnaires de certains types, définis eux-mêmes en termes d'intentions de communication. Par exemple l'acte illocutionnaire de donner un ordre que *p* au moyen de l'énoncé d'une phrase *s* dirigée vers un auditeur *A* sera défini en remplaçant (1) par

(1') *S* a l'intention que *s* produira chez *A* la réponse qui consiste à faire en sorte que *p*.

La notion de convention utilisée par la plupart des auteurs employant les définitions gricéennes est celle de David Lewis (1969) : une convention dans une population *P* est une certaine régularité *R* telle que :

(1) tout le monde dans *P* se conforme à *R* ; (2) tout le monde dans *P* croit que tous les autres en *P* se conforment à *R* ; (3) tout le monde dans *P* a une raison de se conformer à *R*, fournie par sa croyance en (2) ; (4) tout le monde dans *P* préfère en général se conformer à *R* ; (5) il y a une autre régularité possible *R'* qui aurait pu être suivie aussi bien que *R* ; et (6) tout le monde dans *P* connaît (1)-(5) et sait que tout le monde connaît (1)-(5)[1].

On supposera, dans une telle théorie, qu'il y a des conventions spécifiques associées à chaque type d'acte de langage, c'est-à-dire des régularités d'usage pour l'utilisation de phrases d'un certain type pour accomplir les actes correspondants. Il est naturel de supposer d'une part que si *S* *l*-signifie que *p* au moyen d'une phrase *s*, alors il y aura une régularité d'usage d'énonciations de *s* pour *l*-signifier que *p* parmi les membres de la population

1. La notion de « connaissance commune » peut être définie ainsi. C'est une connaissance commune entre *X* et *Y* que *p* si : (1) *X* sait que *p* ; (2) *Y* sait que *p* ; (3) *X* sait que *Y* sait que *p* ; (4) *Y* sait que *X* sait que *p* ; (5) *X* ne croit pas que (4) soit faux ; (6) *Y* ne croit pas que (5) soit faux ; (7) *X* ne croit pas que (6) soit faux ; (8) *Y* ne croit pas que (7) soit faux, et ainsi de suite.

Ces définitions posent des problèmes bien connus, en particulier du fait que la reconnaissance des intentions qui figure dans la définition de la signification du locuteur doit elle-même être commune au sens ci-dessus, en sorte qu'il semble y avoir une régression possible à l'infini des reconnaissances d'intentions. Mais notre but n'étant pas ici de défendre une théorie gricienne des intentions, nous pouvons laisser ces problèmes de côté.

P, et d'autre part qu'il y aura, associées à ces régularités, des conventions proprement linguistiques telles qu'un type de phrase servira à effectuer les actes de *l*-signification correspondants. Ces conventions linguistiques sont, on peut le présumer, indiquées syntaxiquement par les modes : indicatif, impératif, optatif, etc., en sorte que l'on peut supposer que l'indicatif sera à première vue l'indication d'actes d'assertion, l'impératif l'indication d'actes d'ordres, l'interrogatif d'actes de questions, etc.

Peut-on formuler une telle théorie de la force dans le cadre d'un projet comme celui de Davidson ? De prime abord il semble que non. Le programme gricien fait appel à une notion psychologique fondamentale, celle de signification du locuteur, et repose sur l'idée que l'on peut *définir* cette même notion en termes d'intentions. On a dit souvent que le simple recours à la notion d'intention dans une théorie de la signification était incompatible avec la conception vériconditionnelle de la signification, et en ce sens Strawson (1970) a parlé d'un « combat homérique » entre les théoriciens des intentions de communication et les théoriciens des conditions de vérité. Mais ce n'est pas le recours à la notion psychologique d'intention dans une théorie de l'interprétation du discours qui pose problème pour Davidson. Au contraire, comme on l'a vu, une théorie de l'interprétation *doit* faire appel à des notions comme celles des intentions et des croyances des locuteurs. Et si l'on considère que toute énonciation est un certain acte de langage effectué par un locuteur, rien ne nous interdit d'employer ici les concepts de la philosophie de l'action de Davidson, et de considérer que chaque acte d'énonciation peut être *redécrit*, comme toute action, en termes de raisons, c'est-à-dire d'attitudes propositionnelles qu'a un locuteur pour faire cette énonciation. On peut donc très bien concevoir qu'une théorie de la force consiste en une telle redescription, et il n'y a aucune raison d'en exclure les concepts psychologiques[1]. Ce qui fait plutôt problème, pour Davidson, c'est l'aspect réductionniste du programme gricien, l'idée que la signification d'une énonciation isolée, mais aussi la signification linguistique, puisse être définie intégralement en termes d'attitudes psychologiques des locuteurs. Or c'est ce

1. Cf. Davidson, 1975 : 161, *236* : « Une théorie de l'interprétation, comme une théorie de l'action, nous permet de redécrire certains événements de façon à les élucider » ; cf. Mc Dowell, 1976 : 44 ; 1977 : 141 ; 1981 : 119-122.

qu'interdit la méthodologie de l'interprétation : comme on l'a vu, on ne peut pas définir les contenus des attitudes propositionnelles des locuteurs indépendamment et antérieurement par rapport aux significations des phrases qu'ils énoncent. Le programme de Grice reviendrait donc à rejeter ce que nous avons appelé l'interdépendance des croyances (et des autres attitudes propositionnelles) et des significations. Sous certaines de ses versions métaphysiques[1], le programme gricien entre également en conflit direct avec la métaphysique du monisme anomal, puisque les contenus sémantiques devraient pouvoir être réduits à des contenus psychologiques, et ces derniers ultimement à des états physiques. En ce sens, le cadre théorique de Grice est incompatible avec celui de Davidson.

Mais il ne l'est pas si on distingue ce cadre théorique de l'entreprise réductionniste. Il faut en effet distinguer la définition de la signification en termes d'attitudes propositionnelles (a)-(b) d'une analyse de la nature des *raisons empiriques* que peut avoir un interprète d'attribuer des attitudes. Il n'y a rien, dans l'*association* des significations à des intentions qui aille contre la méthodologie de l'interprétation radicale, puisque celle-ci présuppose précisément que l'on ne peut déterminer ce que signifie un locuteur dans une occasion donnée sans déterminer ses croyances et ses intentions. Et il n'y a rien, dans la conception holistique des attributions d'attitudes et de significations, qui aille contre la méthodologie du programme de Grice, puisque ce programme est lui-même fondé sur la dépendance étroite entre les significations et des états psychologiques. Rien n'interdit alors à un partisan du programme davidsonien de prendre les définitions griciennes de la signification du locuteur et de la signification linguistique comme des vérités *conceptuelles* portant sur les relations entre des énonciations de phrases et les états psychologiques des locuteurs, sans préjuger du succès de la thèse supplémentaire selon laquelle ces définitions permettraient de réduire ultimement le sémantique au psychologique. En d'autres termes, nous pouvons nous accorder sur le fait qu'il est impossible d'attribuer des attitudes propositionnelles finement individuées à un locuteur sans être en mesure d'attribuer des significations à ses phrases, sans pour autant rejeter l'idée que des attributions *simultanées* d'attitudes

1. Du moins tel que le présente Schiffer, 1987, par exemple, p. 140 sq.

et de signification soient nécessaires pour évaluer le sens et la force de ses énonciations, c'est-à-dire le fait qu'il conçoive celles-ci comme des assertions, des ordres, ou des questions dotées d'une certain contenu[1]. Mais le problème est maintenant le suivant : peut-on effectuer des attributions simultanées sans recourir à l'autre idée fondamentale incorporée dans le programme gricien, à savoir que les phrases du langage interprété se classent en types d'actes de langage en vertu de *conventions* particulières reconnues par les locuteurs ?

3.3. Modes et conventions

On n'a pas, jusqu'à présent, fait de distinction entre le problème de l'interprétation tel qu'il se pose pour un locuteur isolé (ce que l'on peut appeler l'interprétation *individuelle*) et tel qu'il se pose pour un ensemble de locuteurs ou une communauté (ce que l'on peut appeler l'interprétation *collective*), mais cette distinction nous sera utile, au moins en un sens intuitif. Donner une théorie de la signification pour un certain langage, c'est construire une théorie de la vérité d'une part, et interpréter ce langage d'autre part, c'est-à-dire le décrire, à partir du comportement et des attitudes des locuteurs, comme étant bien *le* langage d'un locuteur ou d'une communauté de locuteurs. Appelons *relation de langage réel* la relation ainsi établie entre un ou des locuteurs et un langage, et concentrons-nous sur le cas collectif, celui où il s'agit d'interpréter le langage d'une communauté ou d'une *population*.[2] Si l'on admet la définition gricéenne de la signification, nous pouvons définir la relation de langage réel d'une population comme l'association conventionnelle à toute phrase d'un langage L d'un type d'énonciation et d'un type de mode spécifique. On supposera que chaque mode agit sur une phrase donnée comme un indicateur de force illocutionnaire, que l'on peut représenter explicitement par des signes comme « $\vdash$ » (pour l'assertion), « ! » (pour l'impératif) ou « ? » (pour l'interrogatif). Ainsi on dira :

(i) *S l-* signifie que $\vdash p$ en énonçant *s* ssi *S* énonce s en ayant l'intention que *A* croie que *p*

(ii) *S l-* signifie que ! *p* ssi *S* énonce *s* en ayant l'intention que *A* fasse en sorte que *p*

pour l'assertion et l'ordre respectivement. On pourrait alors soutenir qu'une langue L est la langue réelle d'une population P si d'une part il y a une fonction qui associe à chaque phrase de L douée d'une signification un certain mode (indicatif, impératif, etc.), et si à chaque mode est associé un certain type d'énonciation (assertion, ordre, etc.) en vertu de conventions dans P d'utiliser des énonciations des phrases à un certain mode pour accomplir des actes de langage du type déterminé par les modes des divers types d'énonciation[1]. De nombreux auteurs ont considéré qu'il existait ce genre d'associations conventionnelles entre des types d'expressions linguistiques et des types d'énonciations. Ainsi Dummett soutient-il que l'assertion est liée conventionnellement à l'objectif de dire ce qui est vrai : de même que dans un jeu le but général est de gagner, dans un langage « la classe des phrases vraies est la classe qu'un locuteur vise à énoncer quand il emploie l'usage assertif de ces phrases »[2]. Il est clair que cette conception des relations entre énonciations et modes est bien trop simple. Il n'y a aucune raison de penser qu'il y ait une régularité d'usage des énonciations d'une phrase comme types d'actes illocutionnaires. Il y a quantité d'actes linguistiques qui ne sont pas des actes de *l*-signification : les parodies, charades, réponses à des examens, répétitions, les compliments volontairement non sincères, etc. Et il n'y a pas nécessairement une relation entre l'emploi d'un mode et un type d'acte de langage. On peut donner un ordre au moyen d'une phrase à l'indicatif, certaines assertions peuvent être énoncées avec un mode interrogatif, etc. On pourrait soutenir qu'il existe d'*autres* conventions permettant de relier des phrases qui ne sont pas, par exemple, au mode interrogatif à l'énon-

1. Cf. Peacocke, 1976 : 107, Evans et Mc Dowell, 1976 : XVI-XVII, Davies, 1981 : 18.

2. Cf. Davidson, 1981 : 267 : « Si nous étions exposés aux locuteurs d'une langue que nous ne connaissions pas, et qu'on nous donne une théorie de la vérité de type tarskien, comment pourrions-nous dire si la définition s'applique à ce langage ? » Dans un vocabulaire plus traditionnel, nous pouvons dire que c'est cette relation qui établit le lien entre la sémantique et la pragmatique. Cf. Peacocke, 1976, Davies, 1981, chapitre 1, Laurier, 1985

1. Cf. Davies, 1981, chapitre 1, pour diverses formulations de cette relation.

2. Dummett, 1973 : 320, et 1976 : 313.

ciation de questions. Mais on ne voit pas comment les spécifier. Il y a deux sortes de façons de résoudre cette difficulté. On peut, tout d'abord, comme Grice, conserver la notion initiale de convention, et modifier la définition de la signification du locuteur, pour rendre compte des cas dans lesquels ce qu'un locuteur *l*-signifie n'est pas ce que *e*-signifient « conventionnellement » les mots qu'il emploie[1]. L'autre solution consiste à proposer une analyse sémantique des modes qui fasse l'économie de la notion de convention. C'est ce que fait Davidson (1979, 1981).

Si le fait d'effectuer un certain acte de langage avec une certaine force est le produit d'une convention, cette convention doit pouvoir être, dans chaque cas, spécifiée. Mais le problème est que l'on ne peut pas préciser laquelle. Le problème classique posé par la notion de convention est que ou bien les conventions en question sont explicites, mais il faut alors les révéler, ou bien elles sont implicites, mais en ce cas on voit mal en quoi le recours à cette notion peut expliquer quoi que ce soit[2]. Supposons que les conventions gouvernant par exemple l'assertion soient explicites. En ce cas, elles peuvent être exprimées par des mots. C'est précisément dans le but de représenter l'acte d'assertion que Frege inventa son signe d'assertion « ⊢ ». Mais il est facile de montrer que l'acte d'assertion ne dépend pas d'une marque conventionnelle comme un signe d'assertion :

> Imaginez ceci : l'acteur joue une scène dans laquelle il est supposé y avoir un incendie (par exemple Tiny Alice d'Albee). C'est son rôle que d'imiter de manière aussi persuasive que possible un homme qui essaie d'avertir les autres qu'il y a un feu. « Au feu ! », crie-t-il. Et peut-être ajoute-t-il, au grand dam de l'auteur, « C'est vrai ! Voyez la fumée ! », etc. Et voici qu'un feu réel s'allume et l'acteur essaie vainement d'avertir les vrais spectateurs. « Au feu ! », crie-t-il, « C'est vrai ! Voyez la fumée », etc. Si seulement il avait eu le signe d'assertion de Frege (1981 : 270, *383*).

Selon l'analyse fregéenne, l'acteur utilise la marque conventionnelle de l'assertion sans avoir l'intention d'exprimer une assertion. Cela est reconnu

1. C'est ce que fait Grice notamment avec sa théorie des « implicatures conversationnelles » (1975), cf. plus bas. Sur les modifications de la notion de *l*-signification, cf. Davies, 1981 : 20 sq.
2. Cf. les critiques classiques de Quine, 1936

par les spectateurs du fait qu'ils ont affaire, sur la scène, à des assertions « feintes », où, comme le dirait Frege, il y a élision du signe d'assertion. Mais si l'acteur veut réutiliser le signe d'assertion avec son sens conventionnel pour faire une assertion, il ne le peut qu'en faisant reconnaître ses intentions. Cela montre, selon Davidson, que même si un trait linguistique a une expression conventionnelle (comme l'usage de l'indicatif pour former des assertions), cette expression conventionnelle peut être utilisée pour un autre but que celui pour lequel elle est conventionnellement utilisée. C'est ce trait que Davidson appelle « l'autonomie de la signification linguistique »[1]. Mais autant dire alors que le caractère « conventionnel » de l'expression ne joue aucun rôle, sauf à dire que le *sens* même des mots est conventionnel, ce qui revient à dire simplement que les mots ont un sens littéral, qui est celui qu'ils ont dans le langage. On peut bien dire que le sens est conventionnel, mais cette proposition n'a aucun caractère explicatif. Avant de voir quelles conséquences Davidson en tire pour l'interprétation du discours, voyons comment il analyse sémantiquement les modes.

Une théorie non conventionnaliste des modes doit satisfaire selon Davidson à trois réquisits. 1 / Elle doit préserver les relations entre les phrases indicatives et les phrases correspondantes des autres modes (elle doit par exemple articuler l'élément commun entre « Vous ôterez vos chaussures », « Otez vos chaussures », et « Oterez-vous vos chaussures ? »). 2 / Elle doit assigner un élément de signification présent dans les phrases d'un mode donné mais non dans un autre. 3 / Elle doit être « sémantiquement viable » c'est-à-dire compatible avec une théorie de la vérité. Mais il y a un conflit potentiel entre ces exigences : les deux premières suggèrent que les modes peuvent être représentés comme des opérateurs sur des phrases, alors que la seconde semble impliquer que les opérateurs en question soient vérifonctionnels, ce qui conduirait à donner une valeur de vérité aux phrases impératives par exemple.

Austin soutenait que les performatifs n'ont pas de valeur de vérité ni de conditions de vérité, mais des conditions de « félicité », et cette idée reste présente dans la plupart des théories des performatifs. La manière

1. 1979 : 113, *173* ; 1975 : 164, *241* ; 1981 : 274-275.

habituelle par laquelle les théories vériconditionnelles de la signification (par exemple Lewis, 1969, 1972) tournent cette difficulté consiste à discerner dans toute phrase performative l'énoncé d'une phrase au mode correspondant, puis dans cette dernière phrase à distinguer l'opérateur de force illocutionnaire du « radical de phrase » commun. Par cette série de transformations, un performatif explicite comme « Pars ! » devient : « Je t'ordonne de partir », puis « Fais en sorte que tu partes ». Lewis admet que cette dernière phrase n'a pas de valeur de vérité, mais que le radical de phrase « que tu partes » en a une, en sorte que la sémantique vériconditionnelle s'applique à lui seulement. Mais dans l'une ou l'autre analyse se pose la même difficulté. Chez Austin, le fait que les phrases performatives n'aient pas de conditions de vérité conduit à distinguer la sémantique de ces phrases de leur sémantique usuelle, donc à supposer que dans un usage performatif un verbe aura une signification différente de celle qu'il a dans un usage non performatif (déclaratif, constatif). Dans l'analyse par « paraphrase » de Lewis, on devra également supposer qu'il y a une différence de signification entre les phrases déclaratives et les phrases énoncées à un autre mode (et tombant dans la portée de l'opérateur de mode). Or le problème est ici le même que celui qui se pose pour le discours indirect et les reports d'attitudes propositionnelles : quelle raison avons-nous de croire que les mots, quand ils tombent dans la portée d'un opérateur, qu'il soit de mode ou d'attitude, ont un sens différent de celui qu'ils ont quand ils tombent dans la portée d'un autre opérateur de mode (par exemple indicatif) ? La solution que propose Davidson ici est parallèle à celle qu'il propose dans son analyse paratactique du discours indirect (§ 1.4). Comme Lewis, il emploie la stratégie consistant à paraphraser une phrase avec un mode donné sous la forme d'une phrase distincte tombant dans la portée d'un opérateur de mode. Mais il refuse de traiter ce dernier comme un opérateur de phrase. Il s'agit plutôt d'un « marqueur de mode » (*mood setter*). Cela donne, pour une phrase assertive :

(4) Il pleut (dit par Jean)
(4a) Jean a fait une assertion dont le contenu est donné par mon énonciation suivante. Il pleut.

et pour une phrase impérative :

(5) Pars !
(5a) J'ordonne ceci. Tu pars.

L'effet de la parataxe est de composer une phrase avec un mode donné en deux actes de langage distincts, l'un par l'énoncé d'une phrase indicative, l'autre par l'énoncé d'un mode spécifique. L'opérateur de mode n'opère pas sur une phrase ; il constitue lui-même une phrase, qui est vraie si la phrase indicative constituant le « noyau » (« tu pars ») a une force illocutionnaire impérative. On ne scinde plus une phrase impérative en deux phrases indicatives distinctes, mais en une phrase qui caractérise une autre phrase indicative comme ayant une certaine force illocutionnaire (le marqueur de mode) et cette phrase indicative. Ainsi les deux phrases ont des conditions de vérité : la première est vraie si la phrase indicative a la force spécifiée, la seconde est vraie si l'indicative est vraie. Cette analyse satisfait, selon Davidson, aux trois conditions d'une théorie des modes. Tout d'abord elle indique l'élément commun, le « noyau indicatif », qui a des conditions de vérité. Ensuite elle représente le mode systématiquement par le marqueur de mode. Enfin, elle est vériconditionnelle : la phrase non indicative n'a pas de valeur de vérité, mais chacune des deux phrases qu'elle conjoint en a une.

Je laisserai de côté pour l'instant les objections qu'on pourrait adresser à cette analyse[16]. Sa caractéristique principale est de concentrer

16. Ces objections seraient semblables à celles qu'on peut adresser à la théorie paratactique. Davidson (1979 : 115, *175*) mentionne lui-même une objection de Hintikka : comment rendre compte, selon cette analyse, des interrogatives ? De prime abord, cette analyse nous prescrit de considérer les interrogatives comme ayant la même sémantique que les phrases indicatives « correspondantes », ou, comme le suggère Davidson, qu'une disjonction de l'indicative et de sa négation. Par exemple « Viens-tu ? » recevrait l'analyse suivante : « Je demande ceci. Tu viens ou tu ne viens pas. » Mais il y a toutes sortes de questions qui ne se laissent pas décomposer de manière aussi simple, comme les questions rhétoriques (« Quelle était ta résolution de nouvel an ? »), les questions d'examen (« La *Begriffsschrift*, quelle année ? »), ou encore les questions destinées à exprimer une surprise (« Dupond vient d'être nommé professeur. — Sans blague ? »). Dans ces cas, il y a plus qu'une information véhiculée par la phrase affirmative correspondante, mais un autre acte de langage, qu'on peut analyser comme une implicature gricienne (cf. ci-dessous) (« Tu avais pris la résolution de ne plus fumer, et je te vois en train de le faire » ; « que cet imbécile ait été nommé professeur m'étonne »). Ou encore Davidson devrait soutenir qu'il y a un second opérateur de mode, indiquant, par exemple, l'ironie, comme (pour la question rhétorique) : « Ce que je vais dire est sur le mode ironique. Je demande ceci. Tu avais pris une résolution de nouvel an. » Et en ce cas, on retrouve les difficultés des complétives enchâssées de l'analyse davidsonienne du discours indirect.

l'expression de la force d'une phrase (de sa nature assertive, impérative ou interrogative) dans l'énonciation (métalinguistique) que fait le locuteur *au sujet* de cette phrase, et par conséquent dans l'intention qu'il manifeste d'utiliser la phrase avec cette force. Cela suppose bien qu'il y ait un certain lien conventionnel entre le mode et la force, mais ce lien tient au sens de la phrase, et pas à une connexion conventionnelle entre les intentions des locuteurs et leurs usages des expressions. Il n'y a donc, outre le sens des phrases, pas d'autre principe d'explication de la force que les intentions des locuteurs. Ces intentions sont nécessairement, selon Davidson, *particulières* et liées à l'acte d'énonciation individuel accompli en une occasion donnée. Cela veut dire que l'interprétation, quand il s'agit de savoir quelle est la relation entre un langage et les locuteurs, est d'abord une interprétation *individuelle*, avant d'être une interprétation collective. Nous allons voir que ce trait s'accentue encore plus dans la façon dont Davidson rend compte des usages non littéraux du discours.

3.4. Sens littéral et sens non littéral

L'une des raisons qui nous incitent à admettre la notion d'un sens « normal », « conventionnel » ou « standard » des mots est l'existence des cas dans lesquels ils sont employés, accidentellement ou intentionnellement, dans un sens non standard, comme dans les pataquès, les jeux de mots, les métaphores, ou diverses figures de discours. Nous appelons alors ce sens « non littéral » par opposition au sens « littéral » que les mots ont dans leur emploi « sérieux » ou « régulier ». Mais Davidson nie que cette distinction ait le moindre fondement, et soutient qu'il n'existe pas plus de principes systématiques non permettant de comprendre un sens littéral qu'il n'en existe pour comprendre un sens non littéral[1]. Cela l'amène à rejeter l'idée que l'interprétation du discours, dans la commu-

1. Cf. 1979 : III, *169*, où il critique la célèbre distinction austinienne entre les usages « sérieux » et les usages « parasitiques » du discours. Il y quelque ironie à constater ici que Davidson s'accorde avec Derrida, qui, dans sa fameuse critique d'Austin et sa polémique avec Searle, contestait également cette distinction. Le rapprochement entre Davidson et Derrida sur ce point a été souligné par Wheeler, 1986. J'ai expliqué pourquoi je trouvais ce rapprochement plus « ironique » que « sérieux » dans Engel, 1991.

nication courante, repose en quoi que ce soit sur la connaissance d'un « langage » ou d'une compétence générale de l'interprète, au sens 1 / d'un ensemble de significations systématiques (déterminées compositionnellement), 2 / partagées par le locuteur et l'auditeur, et 3 / apprises et conventionnelles. Comme on vient de le remarquer, il semble surprenant de rejeter l'idée d'un sens littéral dans les situations linguistiques où les mots ne semblent pas avoir leur sens habituel, précisément parce que ces situations semblent indiquer qu'il existe un contraste entre ce que le locuteur a voulu dire, ses intentions, et ce qu'il a dit en fait. C'est sur ce contraste en particulier que repose la conception gricienne des « implicatures conversationnelles » (Grice, 1975), selon laquelle dans le cas d'une énonciation ironique, par exemple, on doit distinguer le sens littéral ou conventionnel des mots énoncés, de ce que le locuteur a voulu « impliquer » ou « impliciter » conversationnellement. Grice soutient que les locuteurs s'appuient alors sur des principes de conversation ou des maximes de rationalité communicationnelle, qu'ils emploient pour inférer ou « calculer » les intentions du locuteur et interpréter ses énoncés. Un contraste similaire, mais distinct, intervient également dans ce que l'on appelle depuis Donnellan (1966) les « usages attributifs » des descriptions définies par opposition à leurs usages « référentiels ». Davidson ne nie pas qu'on puisse, et sans doute même qu'on doive, employer une telle distinction entre sens littéral et sens non littéral pour rendre compte de l'ironie, de la métaphore, ou des usages référentiels des expressions. Mais il nie que pour comprendre ce que signifie un locuteur dans une occasion donnée on ait besoin de supposer qu'ils partagent une compétence linguistique au sens 1 /- 3 /.

Son argument (1986) s'appuie principalement sur l'analyse des pataquès [*malaproprisms*], c'est-à-dire des cas dans lesquels, pour une raison ou pour une autre, un mot est employé pour un autre (« la démonstration de Gödel est très ingénue », « Denys l'Aérophagite »). Son exemple favori est celui de Mrs. Malaprop, dans la pièce de Sheridan, qui parle d'un « joli dérangement des épitaphes ». Il est naturel de supposer ici que Mrs. Malaprop a voulu dire « épithète » au lieu d'« épitaphe », et qu'elle a simplement confondu les deux mots, et nous n'avons aucun mal à comprendre ce qu'elle a voulu dire. Davidson propose de distinguer deux sortes

de « théories » qu'utilise un interprète pour comprendre ce genre d'énonciations : une théorie « initiale » ou « primaire » [*primary*] de ce que les mots du locuteur signifient, et une théorie « seconde » ou « transitoire » [*passing*], exprimant la manière dont il interprète en fait ce que le locuteur a voulu dire. La même distinction doit s'appliquer au locuteur lui-même, pour qu'il y ait communication ou compréhension : sa théorie primaire est constituée par ce qu'il croit être la théorie primaire de l'interprète ou de l'auditeur, alors que sa théorie seconde est celle qu'il a l'intention que l'interprète utilise. Dans le cas du pataquès, la théorie primaire satisfait aux conditions 1 /- 3 / ci-dessus, p. 127 : elle est apprise par avance et systématique, et on peut la représenter par une théorie-T, et on peut en outre supposer qu'elle est conventionnelle. La théorie seconde peut aussi prendre la forme d'une théorie-T, bien qu'elle ne soit pas apprise, ni, par définition, conventionnelle. Pour que la communication réussisse, il est nécessaire et suffisant que locuteur et auditeur partagent une théorie seconde, et par conséquent celle-ci seule satisfait à la condition 2 /. « L'asymptote de l'accord et la compréhension sont atteintes quand les théories secondes coïncident » (1986 : 442). Dans l'exemple mentionné, la théorie primaire de Mrs. Malaprop *et* sa théorie seconde sont que « un joli dérangement des épitaphes » signifie « un joli dérangement des épithètes » (les deux théories coïncident, parce qu'elle fait une erreur sur le sens des mots tels que devrait le comprendre, selon elle, l'auditeur). La théorie primaire de l'auditeur est que « un joli dérangement des épitaphes » signifie « un joli dérangement des épitaphes », qui est le sens qu'*il* donne à ces mots avant d'interpréter ceux de Mrs. Malaprop, et sa théorie seconde s'accorde avec celle de Mrs. Malaprop. On voit donc qu'en aucun sens la communication ou le succès de l'interprétation ne peuvent reposer sur la capacité à comprendre un « langage » au sens 1 /- 3/. Il est toujours possible de parler ici d'une aptitude linguistique, au sens de « l'aptitude à converger sur une théorie seconde de temps en temps », mais la possession de cette aptitude ne consiste pas en « des règles qu'on utiliserait pour parvenir à des théories secondes, par opposition à des maximes approximatives et des généralisations méthodiques » mais est affaire de « jugeotte, de chance, et de sagesse empruntée à un vocabulaire et une grammaire privée, de connaissance des manières dont les gens s'expriment, et de règles

grossières permettant d'imaginer les déviations à partir du dictionnaire les plus probables » (1986, 446). En ce sens, il est inapproprié de parler de « théories ». Tout ceci, selon Davidson, ne se produit pas seulement dans les usages « non standard » du discours, comme les pataquès, les jeux de mots, les contrepets ou les mots valises, mais dans tous les usages supposés réguliers de la conversation. Il en conclut

> « qu'il n'y a rien de tel qu'un langage, au sens où un langage serait ce que les philosophes et les linguistes ont supposé qu'il en existait, [et]... que nous devons abandonner l'idée qu'il y aurait une structure partagée clairement définie que les utilisateurs du langage acquerraient et appliqueraient dans des cas particuliers et... qu'on doit abandonner la tentative d'élucider la communication en faisant appel à des conventions » (1986, 446).

Cette conclusion peut paraître surprenante, et Davidson soutient qu'elle menace les descriptions usuelles de la compétence linguistique, y compris celles dont il est lui-même responsable (1986, 437). Comment en effet une théorie de la signification pourrait-elle être une théorie de la compétence sémantique d'un locuteur en général, s'il n'existe aucun « langage » dont elle puisse être la théorie ? Certains critiques de Davidson, comme Hacking (1986), y ont vu un renoncement à ses principales thèses en philosophie du langage. Mais il n'y a en fait ici aucun revirement[1]. Davidson continue à soutenir qu'une théorie primaire ou antécédente est une théorie permettant au locuteur ou à l'interprète de comprendre le rôle sémantique d'une infinité de phrases composées à partir d'un ensemble fini de mots, et que cette théorie peut prendre la forme d'une théorie-T. De même une théorie seconde peut prendre cette forme. Mais cela n'implique pas que la compréhension d'une énonciation ou d'un ensemble d'énonciations dans des circonstances données repose sur une telle compétence. L'accord — ou plutôt la compréhension — sont obtenus uniquement à partir de maximes globales comme le principe de charité et de généralisations approximatives sur la façon dont on peut l'appliquer. Les critiques de Davidson disent souvent que le principe de charité et l'usage d'une

1. Bien qu'il soit sans doute vrai que, dans ses premiers écrits, Davidson prenne la notion de « langage », en tant que ce que comprend un locuteur ou ce sur quoi porte une théorie de la vérité, pour acquis.

théorie-T ne peuvent pas suffire pour l'interprétation, et c'est ce qui motive en général le recours à des règles d'interprétation distinctes, comme le principe d'« humanité » ou le principe de projection (§ 2.3). Mais Davidson ne soutient pas que le principe de charité et la théorie systématique que met en œuvre selon lui l'interprète soient des conditions suffisantes. Il soutient seulement que ce principe a pour but de minimiser l'erreur inexplicable, ce qui suppose que l'interprète dispose d'explications autonomes de ce que peuvent être les croyances, les désirs et les intentions plausibles d'un locuteur dans une situation donnée. Ces explications ne peuvent pas elles-mêmes être systématiques ni reposer sur des régularités mentales ou de comportement, sans quoi la théorie de l'interprétation serait elle-même aussi systématique que peut l'être la théorie-T mise en œuvre par l'interprète, et le mental ne serait pas anomal. Pour employer une distinction que Davidson utilise dans un autre contexte, mais à des fins voisines, nous devons distinguer l'explication *à l'intérieur* de la théorie, qui repose sur la structure des phrases et leurs conditions de vérité, telles que les conçoivent interprètes et locuteurs, de l'explication *de* la théorie, qui ne peut être que liée à « des fins et activités humaines » (1977 : 221, *320*, cf. § 6.1). Mais de ces fins et activités nous ne pouvons avoir qu'une conception idéalisée, normative, et non pas les utiliser comme un ensemble de lois du comportement qui seraient vraies ou fausses des individus. On peut voir ainsi l'origine de la confusion que commettent les critiques du principe de charité : supposant à tort que l'interprétation des significations et l'interprétation des croyances doivent se fonder sur des régularités ou des lois strictes, ils se demandent comment le principe de charité ou une théorie systématique de la signification peuvent suffire à l'interprétation, et constatant que l'erreur, l'irrationalité, ou les déviations par rapport aux normes rationnelles d'interprétation sont la règle plutôt que l'exception dans le comportement humain, ils en concluent que le principe de charité est une idéalisation inutile ou gratuite, et que d'autres principes psychologiques doivent entrer en jeu[1]. Mais ni le principe de charité ni une théorie-T ne sont des régularités *descriptives* que

1. Pour d'autres analyses de ce paralogisme au sujet notamment du comportement inférentiel et de l'interprétation des croyances en ethnologie, cf. Engel, 1989, chapitre 13, Engel, 1989a et Engel (à paraître a).

nous puissions attribuer à la psychologie des locuteurs, sous forme de lois du comportement interne à son esprit :

> Les affirmations portant sur ce qui constituerait une théorie satisfaisante ne sont pas... des affirmations portant sur la connaissance propositionnelle d'un interprète, ni des affirmations portant sur les détails du fonctionnement interne d'une partie du cerveau. Ce sont plutôt des affirmations générales quant à ce qu'on peut considérer comme une description satisfaisante de la compétence de l'interprète. *Nous* ne pouvons pas décrire ce qu'un interprète peut faire si ce n'est en faisant appel à une théorie récursive d'une certaine sorte. On n'ajoute rien à cette thèse en disant que si la théorie décrit correctement la compétence d'un interprète, un certain mécanisme dans l'interprète doit correspondre à la théorie (1986, 438).

Ce principe et cette théorie sont des régularités normatives, portant sur ce que *pourrait* utiliser un interprète. Le fait de soutenir, comme le fait Davidson, qu'il n'y a pas d'*autres* principes de l'interprétation, signifie qu'il n'entend pas faire porter à ces principes le poids de l'explication du sens de telle ou telle énonciation. Ce poids doit au contraire être porté, dans le contexte de la communication, par les règles approximatives, les maximes transitoires, ou les indices que l'interprète peut recueillir dans la situation sans pouvoir espérer les généraliser. Mais ces règles n'ont rien de systématique, et par conséquent il n'existe pas de *lois pragmatiques* d'interprétation comme telles[1]. On peut voir ici un parallèle étroit avec la philosophie de l'action de Davidson : les explications du comportement ne peuvent être que des explications causales *singulières*, qui, si elles font appel à des régularités pour prédire ou expliquer, sont toujours susceptibles d'exceptions et non généralisables sous formes de lois intentionnelles. La situation n'est pas différente pour l'interprétation du discours, qui est également une redescription du comportement verbal en termes sémantiques et intentionnels sans que ces descriptions puissent acquérir le type de systématicité qu'on peut attendre d'une science, et sans qu'on puisse les séparer des normes qui les guident[2].

1. Ceci ne veut pas dire que selon Davidson il n'y ait pas de règles pragmatiques qu'on puisse répertorier, mais que ces règles ne peuvent pas être représentées dans un système de lois, ou dans une « logique ». Il n'y a pas pour lui, en ce sens, de « logique illocutoire ».

2. Cf. par exemple Davidson, 1990 : 309 : « Bien que nous puissions quelquefois dire qu'un groupe parle d'une voix, les énonciations sont essentiellement personnelles ; chaque énonciation a son agent et son temps. Une énonciation est un événement d'un certain type, une action intentionnelle. »

On peut pourtant se demander si ces conclusions ne proviennent pas d'une généralisation indue du cas des pataquès ou d'autres usages « non-standard » du discours à l'ensemble des contextes de communication. La description donnée par Davidson de la façon dont nous comprenons Mrs. Malaprop est-elle d'ailleurs la bonne ? Quelle raison avons-nous de supposer qu'elle croit que « épitaphe » signifie « épithète » quand elle énonce la première expression, et par conséquent qu'elle utilise les mots en un sens « non standard » ? Pourquoi ne pas dire ici qu'elle emploie les mots en leur sens littéral — c'est-à-dire en croyant que « épitaphe » signifie « épitaphe » et implique conversationnellement que ce mot signifie « épithète », selon l'analyse gricienne traditionnelle qui distingue ici sens linguistique conventionnel et sens du locuteur[1] ? On peut d'autant plus se poser cette question que Davidson semble admettre tout à fait cette distinction quand il s'agit des usages attributifs et référentiels des descriptions et dans son analyse de la métaphore (1978). Selon lui, les métaphores ne reposent pas sur l'emploi des mots en un sens spécial « métaphorique », fondé sur une comparaison ou une similitude, qui se distinguerait de leur usage « standard » ou « littéral ». Dans les métaphores, les mots ne signifient, rien d'autre que ce qu'ils signifient dans leur sens littéral, mais les intentions du locuteur ne coïncident pas avec ce qui est littéralement énoncé, puisque les métaphores sont littéralement fausses. Qu'est-ce qui distingue cette analyse de celle de Grice, où le locuteur viole la « maxime de qualité » selon laquelle on doit essayer de dire ce qui est vrai ou ce qu'on a de bonnes raisons d'affirmer, et où l'auditeur en infère que le locuteur doit avoir des intentions de signification spécifiques[2] ? L'analyse, comme on l'a vu, fait appel à la notion d'un sens conventionnel, dont on voit mal en quoi Davidson pourrait s'en passer. Il semble y avoir ici la source d'un dilemme. En effet, *ou bien* Davidson renonce à cette notion de sens conventionnel, mais en ce cas il doit finalement la réduire à ce que les mots signifient, à ce qu'ils signifient en une occasion particulière, en fonction des *seules* intentions des locuteurs (qu'il s'agisse de leurs intentions « primaires » ou de leurs intentions

1. Comme le fait remarquer W. Child (1987).
2. J'ai analysé la théorie davidsonienne de la métaphore dans Engel, 1988.

« secondes ») et par conséquent rejeter toute théorie de la signification indépendante des intentions des locuteurs, *ou bien* Davidson l'accepte et doit revenir à l'analyse gricienne traditionnelle. Dummett (1986) a précisément critiqué Davidson sur ce point. Lui attribuant la première branche de l'alternative, il l'accuse de vouloir « libérer les locuteurs de toute responsabilité par rapport au langage en tant qu'institution sociale » et de défendre une forme absurde d'« Humpty-Dumptisme », selon laquelle le sens des mots n'est *jamais* conventionnel et ne dépend *que* des intentions ou des décisions particulières des locuteurs. Comment, dans ces conditions, pourrait-on espérer la moindre convergence dans l'interprétation ? Et comment pourrait-on rendre compte du fait qu'il arrive aux locuteurs de ne pas parler correctement et de faire des erreurs, s'il n'existe pas quelque chose en vertu de quoi on peut parler « correctement » ? Selon Dummett au contraire, on ne peut comprendre les intentions des locuteurs que sur l'arrière-plan d'une « pratique commune » de ceux qui parlent le langage, reposant sur un certain nombre de conventions. Nous devons, en ce sens, employer la notion fondamentale de langage comme ensemble de pratiques linguistiques et de conventions partagées par les locuteurs. L'erreur de Davidson, selon Dummett, est ici de confondre systématiquement compréhension et interprétation, et d'ignorer le point fondamental qu'avançait Wittgenstein quand il disait qu'« il y a une façon de saisir une règle qui n'est *pas* une *interprétation* mais qui est manifesté dans ce que nous appelons « obéir à la règle » dans des cas réels » (1953 : § 201).

On aura l'occasion de revenir sur ce dernier point, et sur les autres aspects de l'opposition entre Davidson et Dummett dans les chapitres qui suivent, mais on peut prévoir aisément la réponse que ferait Davidson sur les autres points. Il soutiendrait sans doute que la question de savoir si Mrs. Malaprop a l'intention de (*l*-) signifier « épithète » quand elle dit « épitaphe » ou si elle croit que « épitaphe » signifie (dans la langue, ou dans la langue de son interprète) « épithète » est indifférente. La première hypothèse consiste à interpréter ses intentions ou ses attitudes à partir de la signification des mots, alors que la seconde consiste à interpréter la signification à partir des attitudes. Mais dans la mesure où attitudes et significations sont pour lui interdépendantes, il n'y a pas, dans l'absolu,

de prééminence d'une hypothèse par rapport à l'autre. Il ne s'ensuit pas que l'une ou l'autre des hypothèses ne soit pas correcte : seul un certain nombre d'autres attributions d'attitudes et de significations à Mrs. Malaprop peut en décider (elle peut par exemple confirmer la première en disant : « Oui, c'est bien ce que je voulais dire » et la seconde hypothèse en disant : « Suis-je bête ! Je ne savais pas que ce que j'appelle « épitaphe » se dit « épithète » », et l'interprète peut confirmer aussi cette hypothèse en l'entendant dire en d'autres circonstances : « L'épithète est ce qu'on écrit sur les pierres tombales »). Il peut y avoir des cas dans lesquels les deux hypothèses sont également plausibles et l'interprétation indéterminée. Mais Davidson soutient qu'il ne s'ensuit pas qu'elle le soit toujours. Les mêmes remarques peuvent s'appliquer au cas des métaphores. Ce qui caractérise les métaphores « vives » par opposition aux métaphores « mortes », c'est que les premières, à la différence des secondes, ne soient pas paraphrasables en termes d'un sens littéral, ou même d'un sens fixant une ou des intentions précises du locuteur ou du poète (« En disant que l'Eglise était un hippopotame il a voulu dire que l'Eglise était un être massif et immobile »). Mais cette différence n'est que de degré : le fait que dans les métaphores, la série des inférences que l'on peut faire quant aux intentions de l'auteur de la métaphore soit indéfinie indique qu'il n'existe pas de limite précise où l'on puisse placer la distinction entre un sens plus ou moins conventionnel et des intentions « créatrices ». A cet égard, Davidson ne refuse pas, comme on l'a vu, de dire qu'en un sens large la notion de signification conventionnelle puisse être utile et même nécessaire, mais ce qu'il rejette est l'idée qu'on puisse faire un contraste net entre ce qui est conventionnel et ce qui ne l'est pas, et entre ce qui relève d'intentions ou d'attitudes psychologiques et ce qui relève du sens des mots[1]. Il ne rejette pas plus la notion de sens littéral — et en ce sens il n'est pas coupable d'Humpty-Dumptisme — mais il réduit

1. Les mêmes remarques pourraient s'appliquer à une conception néo-gricienne comme celle de Sperber et Wilson (1986) qui réduisent les maximes conversationnelles de Grice à la seule maxime de pertinence. Mais si l'on peut s'accorder avec Sperber et Wilson pour dire que le principe de pertinence est la seule règle générale de communication, elle a ceci de commun avec la charité selon Davidson qu'il est impossible de fixer son contenu si elle est une règle normative. Et si elle est, comme ils le soutiennent, une propriété cognitive de la psychologie des agents humains, il est encore plus difficile de lui donner un contenu.

cette notion aux seules *conditions de vérité* des phrases (1981 : 269). Il ne s'ensuit pas non plus qu'il rejette les descriptions que l'on peut donner dans un cadre gricien des actes de langage indirect, de l'ironie, des tropes, etc., puisque, comme on l'a vu, ses analyses sont en fait très proches sur ces points de celles de Grice. Il ne s'agit pas pour Davidson de bannir de la description de ces phénomènes le vocabulaire psychologique ou « cognitif » des intentions, des attitudes, des inférences. Ce dont il doute, en revanche, c'est de la validité du programme gricien en tant que tentative de *réduction* de la signification à des attitudes psychologiques.

Les mêmes remarques s'appliqueraient à toute description de la communication en termes d'*idiolectes*. On pourrait en effet soutenir que l'interprétation individuelle que Davidson décrit montre que les locuteurs ne partagent pas un langage commun, mais qu'ils emploient des idiolectes, entendus comme des ensembles de significations qu'ils sont les seuls à avoir[1]. En ce sens « épitaphe », dans l'idiolecte de Mrs. Malaprop, signifie « épithète », et les différents usages non standard correspondent à des emplois idiolectaux, transitoires ou permanents. La théorie « primaire » de chaque interprète serait celle de son idiolecte. Mais la notion d'idiolecte n'est pas plus claire que celle de « langage », y compris quand ce terme renvoie à des sous-langages de langages parlés par une population[2]. Du point de vue davidsonien, qu'est-ce qui nous permettrait de distinguer le cas où Mrs. Malaprop a ces significations dans son idiolecte, d'un cas où elle emploie un idiolecte différent (un sous-idiolecte du sien ?), ou d'un cas où ses intentions de communication divergent des sens de son idiolecte ? Il n'y a pas de fondement à ces distinctions. Dans la perspective de Dummett, l'idiolecte d'une personne correspond à des croyances imparfaites ou erronées portant sur les significations d'un langage d'une communauté. Il n'y a pas, en ce sens, de distinction à faire entre la grammaire et la sémantique du langage commun et celles de l'idiolecte : c'est seulement le locuteur qui fait des erreurs sur les sens conventionnels des mots[3]. A nouveau, comment l'erreur serait-elle

1. Davidson semble vouloir employer cette notion quand il écrit : « La théorie primaire possède tous les traits spécifiques à l'idiolecte du locuteur » (1986 : 443).
2. Cf. les remarques de Chomsky, 1986, et de George, 1990.
3. Cf. Dummett, 1975 : 135, 200 ; 1978 : 425.

possible si l'on ne supposait pas que quelque chose est la norme ? Mais s'il y a des cas dans lesquels il est clair qu'un locuteur diverge involontairement de la norme, comment le déterminer dans tous les cas (le député qui pendant la cérémonie d'intronisation du roi crie « Vive la République ! » peut confondre les institutions, faire un pataquès non intentionnel, faire « intentionnellement » une erreur, avoir l'intention de prononcer les mots en leur sens littéral et de faire reconnaître cette intention aux auditeurs, etc.)[1] ? Mais une chose est de dire que la plupart des croyances sur lesquelles porte l'interprétation doivent être vraies pour que l'erreur soit possible, autre chose est de dire que l'on peut déterminer exactement quand les locuteurs se trompent sur le sens des mots ou les emploient de façon non conventionnelle, et quand leurs croyances sont erronées plutôt que leurs significations. Le seul moyen, selon Davidson, de parvenir à désintriquer les rôles respectifs des intentions et des attitudes par rapport à celui des significations dans l'interprétation est de partir d'une attitude non individuative, celle de « tenir-pour-vrai » ou de « préférer-vrai » et d'élaborer, sur la base d'une théorie primaire, une théorie seconde des énonciations d'un locuteur identifiant la part respective des croyances et des significations. Mais il n'y a aucune raison de penser qu'une telle théorie soit autre que transitoire, c'est-à-dire qu'elle puisse fixer ce que nous pourrions appeler un idiolecte d'un locuteur, et *a fortiori* le langage d'une communauté de locuteurs. Si c'est le cas, alors le projet évoqué ci-dessus de formuler les conditions nécessaires et suffisantes pour qu'un langage soit le « langage réel » d'une population est nécessairement voué à l'échec, si par « langage » on entend quelque chose qui satisfait aux conditions 1 /- 3 / ci-dessus, p. 127. Rien n'interdit en revanche, si l'on entend par « langage » *ce à quoi s'applique une théorie de l'interprétation*, de soutenir qu'une théorie-T pour le langage d'une population *P* est le langage réel de cette population si elle maximise le nombre de phrases tenues-pour vraies par ses locuteurs qui sont effectivement vraies

1. Ces remarques s'appliqueraient aux cas, très discutés, dans lesquels un locuteur emploie deux nom propres distincts pour désigner un même lieu distinct (comme le Pierre de Kripke (1979) avec « Londres » et « London ») ou un terme d'espèce naturelle dans un sens déviant (« arthrite » selon Burge, 1979). Je ne discuterai pas ici ces cas. Cf. Bilgrami, 1992, qui développe en détail une analyse de ces cas. Cf. également ci-dessous, § 6.5, note 43.

d'après les conditions de vérité spécifiées par T, moyennant les autres contraintes de l'interprétation.

3.5. Dire vrai et tenir-pour-vrai

Nous pouvons essayer d'envisager les problèmes traités dans ce chapitre sous un autre angle, en considérant à nouveau l'attitude que Davidson considère comme le pivot de sa théorie de l'interprétation, celle de tenir-pour-vraie (ou préférer-vraie) une phrase. On pourrait faire les suggestions suivantes. Pourquoi l'interprétation devrait-elle partir de cette attitude, et non pas d'un type d'énonciations particulières que nous appelons des *assertions* ? Non seulement asserter que *p* semble être l'expression naturelle du tenir-pour vrai, mais cette notion semble étroitement liée à celle de vérité : qu'y a-t-il de plus dans la notion de vérité que dans celle d'assertion ? Qu'y a-t-il de plus dans « '*p*' est vrai » que l'affirmation ou l'assertion que *p* ? Par conséquent, comment pourrait-on chercher à élucider la notion de signification en faisant appel à celle de vérité sans recourir au concept d'assertion ? Ces suggestions peuvent venir aussi bien de théoriciens qui, comme Grice et Strawson, soutiennent que l'on doit analyser la notion de signification en termes d'intentions de communication (asserter que *p* c'est avoir l'intention de communiquer que *p*)[1], que de théoriciens qui, comme Dummett, soutiennent que l'on ne peut comprendre ce que c'est que comprendre et parler un langage sans faire appel à l'idée d'une convention générale selon laquelle les locuteurs *visent* à dire la vérité, et donc à asserter des phrases.

La réponse de Davidson est prévisible : partir de la notion d'intention de communiquer que *p*, c'est supposer que l'on sache quels sont les contenus des intentions avant même de savoir ce que les agents signifient par leurs mots. Et partir de la notion d'assertion en supposant qu'elle est liée à une convention générale d'après laquelle les locuteurs visent à dire la vérité, c'est fonder une notion obscure sur une autre notion obscure. Or nous avons vu qu'il n'y a pas, selon Davidson, en raison de

1. Cf. par exemple Strawson, 1970 : 180.

l'autonomie de la signification linguistique, de convention générale de ce genre, et pas de marques conventionnelles de l'assertion. Nous ne pouvons donc pas fonder la théorie de l'interprétation sur la notion d'assertion, ni sur la reconnaissance par l'interprète d'actes d'assertion. La notion d'assertion est beaucoup plus riche que celle de tenir-pour-vraie une phrase. Quelqu'un qui asserte une phrase doit se représenter comme tenant cette phrase pour vraie, ou comme la croyant vraie, et comme l'énonçant. Selon l'analyse des modes indiquée plus haut, celui qui asserte qu'il pleut se représente comme ayant fait une assertion dont le contenu est donné par son énonciation de la phrase « Il pleut ». Et si nous voulons dire que quelqu'un a fait une assertion en énonçant « Il pleut » nous devons le représenter comme ayant fait une assertion dont le contenu est donné par *notre* énonciation de la phrase « il pleut », et nous représenter, selon l'analyse parataxique, comme « mêmediseurs » que lui. Mais cela suppose que nous connaissions le sens de la phrase « Il pleut », ce que nous ne pouvons pas présupposer dans l'interprétation radicale. En d'autres termes, asserter que *p* suppose que le contenu de *p* soit déjà individualisé, et si nous voulons dire qu'asserter que *p* est la même chose que dire que *p* est vrai, cela suppose que nous sachions *quelle* vérité *p* exprime (ou quelle proposition). C'est pourquoi l'interprétation doit se fonder sur une notion moins riche, celle de tenir-pour-vrai.

Ce point se rattache étroitement à une question sur laquelle nous reviendrons dans les chapitres suivants, celle de la définition de la notion de vérité. Supposons que nous demandions quelle notion de vérité doit être utilisée par une théorie de la signification du type de celle que propose Davidson. (Il ne s'agit pas ici de se demander quelle conception de la vérité est la bonne, mais quelle conception de la vérité est la bonne *relativement* au projet d'éclairer la signification au moyen de la notion de vérité.) Considérons d'abord la thèse évoquée ci-dessus selon laquelle il n'y a rien de plus dans la notion de vérité que l'équivalence entre « il est vrai que *p* » et « *p* ». Elle est connue sous le nom de théorie de la vérité-redondance[1]. Mais puisque cette équivalence revient à l'équivalence

1. Cf. Dummett, « Truth » (1959), *in* Dummett, 1978, Horwich, 1990, Engel, 1989, chapitre V. Davidson discute les différentes conceptions de la vérité dans 1990.

entre « il est vrai que *p* » et l'*assertion* de p, et que l'assertion de *p* suppose que nous ayons individualisé le contenu de *p*, la théorie de la vérité-redondance présuppose une connaissance des significations[1]. Elle ne peut donc être compatible avec la conception davidsonienne de l'interprétation radicale. On pourrait ensuite envisager qu'il n'y ait rien de plus dans la notion de vérité que le schéma tarskien « '*s*' est vrai ssi *p* ». Ce schéma n'est pas équivalent au précédent, qui prédique la vérité de contenus de croyances, ou de propositions, alors que celui-ci prédique la vérité de *phrases*. Selon cette conception, la fonction du prédicat « vrai » est seulement de nous permettre de « déciter » la phrase entre guillemets à gauche du biconditionnel, et c'est pourquoi on appelle cette conception « décitationnelle »[2]. En apparence, c'est précisément sur cette conception décitationnelle que repose la théorie davidsonienne de la signification, puisqu'une théorie-T prend la forme d'un ensemble de phrases-T conformes au schéma décitationnel. Mais cette apparence est trompeuse. D'une part le schéma décitationnel ne peut définir la vérité qu'en présupposant la notion de traduction[3], et d'autre part il ne peut pas définir la *notion* de vérité en général, indépendamment d'*un* langage *L* particulier, et par conséquent ne peut définir le prédicat « vrai » comme un prédicat commun à un certain nombre de langages. Comme on l'a vu, une théorie-T ne nous donne que l'*extension* du prédicat de vérité, et ne le définit pas[4]. La solution de Davidson, comme nous le savons, consiste à admettre que les théories-T ne définissent pas le concept de vérité, et à prendre le prédicat de vérité comme un prédicat primitif, non défini, pour appliquer ensuite la structure révélée par une théorie-T à l'élucidation de la signification, et à utiliser les contraintes de l'interprétation radicale pour passer des phrases-T tenues vraies à des assignations de signification et de croyances.

Mais cette solution suppose que l'on puisse isoler, chez les locuteurs, l'attitude en question (ou celle de préférer-vrai). Nous savons pourquoi

1. Cf. Baldwin, 1991 : 22-25.
2. Cf. Field, 1986, Horwich, 1990, Putnam, 1983, Soames, 1984, Davidson, 1990, Baldwin, 1991.
3. Cf. Davidson, 1990 : 296 ; Baldwin, 1991 : 25-30.
4. Davidson, 1990 : 285-293.

le choix de ces attitudes est requis : c'est parce qu'elles sont « applicables à toutes les phrases » et parce qu'elles « ne nous demandent pas d'être capables de faire des distinctions finement discriminées entre des croyances » (1974 : 135, *200*). Et Davidson ajoute que « c'est une attitude dont on peut supposer de façon plausible que le locuteur est capable d'identifier avant qu'il puisse interpréter, car il peut savoir qu'une personne a l'intention d'exprimer la vérité sans savoir en quoi que ce soit *quelle* vérité est exprimée » (*ibid*). Mais comment identifie-t-on le tenir-pour-vrai (ou le préférer-vrai) ? Si nous ne voulons pas employer des marques purement comportementales, comme le voudrait la stratégie quinienne de repérage de l'« assentiment » et du « dissentiment », le seul moyen semble être de repérer quelque chose comme des marques conventionnelles de « l'intention d'exprimer la vérité », et par conséquent des marques de ce qu'il faut bien appeler des assertions. Davidson dit lui-même, une phrase après le passage cité ci-dessus, qu'il faut détecter les attitudes d'assertion, mais aussi de mensonges, d'ordres, d'ironie, etc., qui peuvent toutes « révéler si un locuteur tient ses phrases comme vraies » (*ibid*). Mais comment détecter ces attitudes, et comment savoir si celles-ci sont bien des attitudes par rapport à la vérité si ce n'est par rapport à divers actes illocutionnaires reconnus comme des assertions, des ordres, etc. ? Il est donc faux de dire que Davidson ne recourt pas à la notion d'assertion pour repérer les marques du tenir-pour-vrai. Il admet d'ailleurs (1981 : 275) que son rejet de l'association de la notion d'assertion à des conventions n'exclut pas « qu'il n'y a de relation entre les indicateurs de modes et l'*idée* d'un certain acte illocutionnaire ». Selon sa théorie parataxique des modes, il soutient qu'une énonciation d'une impérative « *se signale [label]* elle-même comme un ordre ». Quand la syntaxe contient un marqueur de force, ce label est clair, et en ce sens on peut bien parler de convention, mais pas au-delà de l'idée que « la convention est gouvernée par le sens littéral ». Mais que dire dans le cas où cette marque syntaxique n'est pas explicite ? Que le locuteur doit « se représenter lui-même » comme énonçant une assertion un ordre, etc. ? Mais comment peut-il se représenter ainsi s'il ne sait pas ce qu'est une assertion, un ordre, etc. ? En ce sens, les parataxes du type proposé (comme (4a) et (5a) ci-dessus, p. 124-125) présupposent que les locuteurs aient une notion de ce qu'est une assertion.

Il paraît difficile de dire qu'ils puissent avoir des attitudes de tenir-pour-vrai indépendamment d'une conception générale de la relation qui existe entre « l'intention d'exprimer la vérité » et le type d'acte linguistique capable de réaliser cette intention. Il paraît également difficile pour l'interprète de pouvoir attribuer l'attitude de tenir-pour-vraie une phrase dans des contextes d'énonciations ironiques, mensongères, ou fictionnelles sans avoir une connaissance plus ou moins détaillée de la psychologie des agents, qui nous permette de distinguer ces attitudes d'attitudes voisines telles que tenir-pour-probable, ou tenir-pour désirable, etc., c'est-à-dire sans recours à un certain nombre d'attitudes plus individuatives que celle de tenir-pour-vrai [1].

Davidson semble devoir naviguer entre deux écueils : s'il reconnaît que la stratégie de l'interprète doit se fonder sur une notion comme celle d'assertion, il risque de réintroduire la notion de convention qu'il veut éviter, et s'il admet qu'on doit discerner chez les locuteurs d'autres attitudes que celle de tenir-pour-vrai, la thèse de l'interdépendance des croyances et des significations est menacée, et la vérité perd sa place centrale dans une analyse de la signification. La première branche de l'alternative invite à des réactions comme celle de Dummett pour qui il faut supposer une convention générale liant les actes d'assertion à l'objectif général de dire le vrai. La seconde invite à des réactions comme celle de Baldwin :

> Davidson ne nous fournit pas de raison de supposer que le concept de vérité joue un rôle explicatif essentiel dans la transition conduisant des données empiriques concernant les conditions dans lesquelles les locuteurs tiennent des phrases pour vraies à l'interprétation de ces phrases. Au lieu de cela, il est raisonnable de supposer que la transition est accomplie par une théorie interprétative qui combine ces données avec une compréhension de l'information disponible aux locuteurs, les activités dans lesquelles il est plausible qu'ils soient engagés, et les conventions qui informent leurs actes linguistiques. Nous ne pouvons pas simplement aller de l'observation selon laquelle Jacques tient pour vraie la phrase apparemment déclarative *s* à la conclusion que dans la bouche de Jacques *s* signifie que *p* ; nous avons besoin de poser l'hypothèse que Jacques exprime la croyance que *p* sur l'arrière-plan de notre compréhension de ses autres croyances, ou des actions que la croyance que *p* le conduiraient probablement à accomplir, ou des

1. Bennett, 1985 : 612.

phrases apparemment similaires à *s*, et du type d'acte de langage qu'est son énonciation de *s*. Mais nulle part il n'y a de rôle direct pour des questions portant sur la vérité de *s*, par opposition à des attitudes vis-à-vis de sa vérité. La vérité est un concept inapproprié pour ce rôle : l'interprétation repose sur des faits psychologiques et sociaux, et les questions portant sur la vérité sont trop impersonnelles pour être utiles à cette étape. A partir du moment où la signification a été identifiée, elle peut bien entendu être représentée en termes de conditions de vérité,..., mais ce n'est qu'un exercice dérivé de représentation et non d'élucidation» (Baldwin, 1991 : 31-32).

Il y a dans ce que dit Baldwin un certain nombre de choses avec lesquelles Davidson n'a pas besoin d'être en désaccord : l'interdépendance des croyances et des significations, le holisme des croyances, ou le fait que l'interprétation ne prend sens que dans un contexte d'activités humaines sociales. Davidson ne nie pas que l'interprétation et la communication courantes prennent nécessairement place dans ce contexte (qui le nierait ?). Son problème est fondationnel : il est celui de savoir si l'on peut parvenir à une conception philosophiquement informative de la signification à partir de notions comme celles de « convention », de « pratique » ou d'« activités humaines ». Et sa réponse, comme on l'a vu, est négative. *Partir* de ces notions présupposerait ce qui est en question[1]. Mais il ne s'ensuit pas, d'un autre côté, que Davidson cherche à *définir* la notion de signification, au sens d'une réduction de ces notions à des notions plus fondamentales (extensionnelles, ou psychologiques) : cette entreprise lui paraît autant vouée à l'échec que celle qui consiste à réduire l'intentionnalité à des notions plus primitives[2]. Mais comment concilier ces deux

1. Cf. par exemple 1976 : 171, *252* : « Tout comme Lear gagne en pouvoir en l'absence de Cordelia, je pense que les traitements du langage prospèrent quand ils évitent l'évocation non critique des concepts de convention, de règle linguistique, de pratique linguistique, ou de jeux de langage. »

2. Cela permet, à mon sens, de répondre à une objection que m'a adressée Baldwin : si Davidson entend définir la signification en extension, comment peut-il éviter la difficulté (socratique) de toute définition de ce genre, à savoir la question de savoir comment l'appliquer à des cas nouveaux ? A mon sens, la réponse que donne Davidson à Field, que j'examine au § 6.1 ci-dessous, est la bonne. La réponse est que Davidson ne fournit pas une *analyse* ou une définition de la notion de signification, ni en compréhension, ni en extension. Il fournit ce que l'on pourrait appeler une caractérisation de ce concept dans le cadre des conditions de l'interprétation radicale. Il est donc vrai qu'il n'y a pas de critères stricts permettant d'appliquer celle-ci à des cas nouveaux. Le problème est que des notions comme celles de « signification » ou d'« intentionalité » ne répondent

exigences, celle d'une base conceptuelle assez pauvre et la reconnaissance de l'impossibilité d'une réduction en bonne et due forme ? Comme on l'a vu, en essayant de limiter cette base conceptuelle au *minimum* : à des notions comme celles de tenir-pour-vrai, de relations des croyances à des phrases ou inscriptions, et en recourant au principe de charité pour combler la lacune entre ce qui est tenu-pour vrai et ce qui est vrai. Comme de nombreux critiques, Baldwin juge le principe irréaliste et la notion de vérité trop « impersonnelle » pour interpréter les contenus sémantiques et mentaux intervenant dans le cadre d'activités et d'intérêts humains complexes. Il me paraît ici faire l'erreur que j'ai signalée plus haut, qui consiste à prendre un principe normatif pour un principe descriptif.

Le dilemme suggéré au paragraphe précédent est largement artificiel. Malgré certaines de ses déclarations explicites sur ce point, Davidson n'a pas besoin d'évacuer tout concept psychologique ni toute notion de convention de son analyse. Que le sémantique ne se réduise pas au psychologique n'implique pas qu'il ne survienne pas sur le psychologique (au sens de § 2.5), et par conséquent qu'il n'y ait pas de liens entre les deux. On peut admettre qu'il soit impossible de présupposer des états intentionnels « finement discriminés » pour rendre compte des significations linguistiques, sans pour autant se dispenser de certaines attributions d'attitudes *suffisamment* discriminées pour prendre pied dans l'interprétation. Si nous devons, par exemple, essayer de repérer une attitude de tenir-pour-vraie une certaine phrase par un locuteur qui l'énonce de façon ironique, il nous faut des hypothèses sur sa psychologie. L'élargissement de la base psychologique du tenir-pour vrai au préférer vrai, et l'inclusion des désirs dans le cadre bayésien évoqué au § 2.7, montrent qu'on ne peut s'en tenir à une seule attitude des agents. Après tout, l'interprétation, tout comme l'explication de l'action est une certaine forme de rationalisation, et on ne voit pas pourquoi elle ne pourrait pas partir de combinaisons de croyances, de désirs et d'intentions plus complexes que des croyances

(Suite de la note 2, p. 132)

pas à des critères précis. Si la position « minimale » décrite aux chapitres 5 et 6 est correcte, et si elle s'applique, au moins en grande partie, à la position de Davidson, cette position ne nous engage en rien à définir la notion de signification en termes plus primitifs. Mais elle n'implique pas, comme la position que je décris comme « quiétiste », qu'il n'y ait rien à en dire non plus.

et des désirs quant à la vérité de certaines phrases. On ne voit pas non plus pourquoi des contraintes interprétatives comme le principe d'humanité, qui prescrivent de supposer une certaine similarité entre l'univers doxastique de l'interprète et celui de l'interprété ne pourraient pas être utilisées, et j'ai précisément soutenu au § 2.3 que ce principe n'était pas incompatible avec celui de charité. Et si l'interprétation du discours doit reposer sur une théorie « seconde » au sens ci-dessus, on voit mal comment cette théorie pourrait se passer de ce genre d'attributions et de principes[1]. Si l'on acceptait d'élargir ainsi la base empirique de l'interprétation, il ne s'ensuivrait pas pour autant que les attitudes par rapport à la vérité de phrases perdraient leur place centrale dans la théorie de l'interprétation. Même si l'on rejette, comme Davidson, l'idée qu'on puisse *expliquer* la compétence sémantique des locuteurs au moyen de notions comme celles de convention et d'assertion, il me semble indispensable de conserver une place à ces notions dans une théorie de l'interprétation. Sans quoi on ne comprendrait pas pourquoi les attitudes des agents par rapport à la *vérité* sont la dimension principale d'*évaluation sémantique* des phrases. Dummett, à mon sens, a raison sur ce point : c'est parce que la vérité est une dimension première d'évaluation de la signification, et parce qu'elle est visée par les locuteurs qui assertent des phrases, qu'elle occupe une telle place dans une théorie de l'interprétation. Il y a à cet égard une notion *normative* de vérité, au sens où la vérité est *ce qui vaut la peine* d'être recherché ou d'être asserté, distincte de la vérité comme propriété sémantique des phrases, au sens où elles se trouvent, de fait, être vraies, et il me semble que Davidson devrait reconnaître ce fait. Nous pouvons bien appeler cette norme une convention sans supposer qu'elle correspond à une régularité d'usage. En d'autres termes, elle pourrait être violée la plupart du temps sans que cela menace son statut normatif. Nous pouvons désigner cette notion, à la suite de Wiggins (1980), comme étant celle de *Vérité*. Wiggins la justifie ainsi :

1. C'est en ce sens que des auteurs comme Peacocke (1983) et Mc Ginn (1986) ont plaidé pour un élargissement de la base de l'interprétation radicale. Je l'ai défendu également dans Engel, 1988a. Cf. ci-dessous, § 6.5 *in fine*.

> Supposons que l'assertabilité [la *vérité*] ne soit pas la dimension première d'évaluation. Alors il n'y aurait aucune raison pour que ce ne soit pas une *norme* pour les locuteurs d'énoncer seulement des phrases qu'ils tiennent comme des phrases assertables... *A fortiori*, il n'y aurait généralement rien dans la pratique des locuteurs de L pour soutenir l'idée que c'était une norme pour les locuteurs de L que de viser l'assertabilité (ou la vérité) dans leurs énonciations. Mais s'il n'y avait pas une telle norme, s'il n'y avait pas d'attente rationnelle qu'il y ait une telle norme, alors l'interprétation serait impossible... ou injustifiable. Supposons que l'énonciation d'une phrase n'ait aucune connexion systématique avec la croyance du locuteur, et que l'énonciation ne donne aucune raison à première vue à l'interprète d'attendre qu'une croyance a été exprimée. Alors un lien vital serait perdu avec ce qui peut donner à une interprétation un soutien empirique indépendant, à savoir ce qu'il serait *rationnel* à un individu quelconque de croire, s'il était placé comme ce sujet est placé dans cet environnement (1980 : 205-206).

Il me semble indispensable, dans une conception qui insiste sur le caractère « normatif » de l'interprétation, de maintenir l'existence d'une telle dimension de *Vérité*, et d'admettre qu'elle est, comme le dit Dummett, ce vers quoi tendent nos assertions. Si c'est le cas, il y a *plus* dans la notion usuelle de vérité que l'équivalence entre « '*p*' est vrai » et « *p* » et que le schéma décitationnel « '*p*' est vrai ssi *p* »[1]. Il doit donc exister des liens manifestables entre cette dimension normative de la vérité et la pratique de l'interprétation. Wiggins propose d'appeler ceci une « marque de la vérité » permettant de relier le concept de vérité (T), tel qu'il figure dans une théorie-T, au contexte général de l'interprétation, et la formule ainsi :

> (i) Le désir des locuteurs de donner leur assentiment à une phrase doit, en général, être une indication de leur croyance qu'elle est T ; T doit être une propriété que les assertions doivent être normalement interprétés comme visant (*ibid*).

Le concept de *Vérité* se distingue donc du concept de vérité, en ceci qu'il contient une force normative le reliant à ce que nous sommes supposés faire quand nous assertons des phrases. Mais cela ne nous dit pas encore quel est le lien exact entre les deux concepts. Répondre à cette question

1. Je reviens sur ce point au chapitre 5 ci-dessous.

n'est pas facile : cela suppose que nous sachions ce que signifie le concept de vérité que nous avons employé dans la formulation d'une théorie de la signification en termes de « conditions de vérité ». C'est à cette question que nous devons nous adresser à présent.

4

Le défi antiréaliste

> Comprendre une proposition, c'est savoir ce qui est le cas quand elle est vraie. (On peut donc la comprendre sans savoir qu'elle est vraie.) On la comprend quand on comprend ses constituants.
>
> L. Wittgenstein, *Tractatus*, 4024.

4.1. Réalisme et antiréalisme en théorie de la signification

On sait que le slogan « le sens c'est l'usage » n'est chez Wittgenstein qu'une recommandation, destinée à nous mettre en garde contre la tentation de réduire la signification à un seul aspect central, ou d'entreprendre la construction d'une théorie systématique de la signification[1]. Dans le chapitre précédent, nous n'avons envisagé qu'une interprétation de ce slogan d'après laquelle la signification des phrases pourrait être constituée, outre par leurs conditions de vérité, par leurs conditions d'usage dans divers contextes par des locuteurs. Mais il y a une interprétation beaucoup plus radicale du slogan, selon laquelle la signification est, en un certain sens, *identique* à l'usage. Il y a deux versions possibles de cette interprétation. Selon la première, apparemment plus proche de la lettre wittgensteinienne, la signification d'une phrase n'est pas autre chose que la variété de ses usages, réels ou possibles. Cette variété étant indéfinie, il s'ensuit qu'il ne peut pas y avoir de théorie vraiment systématique de la signification, mais seulement des analyses particulières des rôles que des phrases ou des types de phrases peuvent jouer dans nos « jeux de langage » et dans nos « formes de vie ». La seconde version, qui nous occu-

1. Wittgenstein, 1951, § 43 : « Pour une *large* classe de cas — bien que pas tous — dans lesquels nous employons le mot "signification", il peut être défini ainsi : la signification d'un mot est son usage dans le langage. »

pera ici, consiste à isoler un trait particulier de l'usage du langage, distinct des conditions de vérité, et tenu comme central. On peut considérer que ce trait est la capacité qu'ont les locuteurs, quand ils énoncent certaines phrases, de connaître leur signification, et de manifester cette capacité de manière publique. Selon Dummett, qui a proposé cette interprétation du slogan wittgensteinien[1], ce dernier n'entre pas en conflit avec le projet d'une théorie systématique de la signification parce qu'il est possible de déterminer, pour tout type de phrase, la nature spécifique de son « usage ». Mais il entre en conflit avec le projet d'une théorie systématique vériconditionnelle de la signification. Car

1 / à supposer que la signification consiste bien dans les conditions de vérité, et la connaissance de la signification dans la connaissance des conditions de vérité, cette connaissance devra être intégralement manifestable dans l'usage que les locuteurs font des phrases.

2 / Mais il est clair qu'il y a de nombreuses phrases de notre langage (y compris des phrases que nous n'avons jamais rencontrées) dont nous ne connaissons pas les conditions de vérité, bien que nous comprenions ces phrases (ou que nous pourrions comprendre si nous les rencontrions).

3 / Par conséquent la thèse selon laquelle ce que nous comprenons, quand nous comprenons ces phrases, sont leurs conditions de vérité, doit être fausse. Ce que nous comprenons relève plutôt de l'usage de ces phrases.

C'est à l'élucidation de cet argument — que nous appellerons « argument antiréaliste standard » — qu'est consacré ce chapitre. Sa force vient de ce qu'il procède d'une prémisse identique à celle de Davidson : le rôle d'une théorie systématique de la signification est de nous fournir une certaine analyse de ce que connaissent les locuteurs quand ils connaissent leur langage. Mais Dummett soutient qu'une théorie vériconditionnelle de la signification est incapable de représenter correctement cette connaissance, parce qu'elle présuppose que si la signification consiste dans les conditions de vérité, les locuteurs n'ont pas besoin de reconnaître ces conditions de vérité, qui peuvent donc transcender les moyens qu'ils ont de les saisir. Cette hypothèse est selon Dummett indissociable d'une conception

1. Dummett, 1978 : 443-444 ; 1975 : 101 ; 197 : 72, 135.

réaliste de la signification : les conditions de vérité, et par conséquent la signification, des phrases sont indépendantes de la connaissance que nous pouvons en avoir. Elle est également constitutive du réalisme en général, comme thèse métaphysique. L'argument que Dummett dirige contre la conception vériconditionnelle est donc étroitement lié à celui qu'il dirige contre cette cible métaphysique plus générale. Mc Ginn (1980) a bien caractérisé cet argument comme une contraposition ou un *modus tollens*[1].

(a) Si le réalisme est vrai, alors la signification transcende l'usage ;
(b) mais la signification ne transcende pas l'usage ;
(c) donc le réalisme est faux.

La prémisse (a) est l'interprétation que donne Dummett de la thèse réaliste ; la prémisse (b) est l'interprétation spécifique qu'il donne du slogan wittgensteinien ; la conclusion (c) revient à l'affirmation de la thèse opposée au réalisme, que Dummett appelle *antiréalisme*. La portée de cet argument est très générale : pour Dummett le réalisme est, avant d'être une thèse métaphysique sur la nature de la réalité ou une thèse épistémologique sur la connaissance que nous en avons, une thèse *sémantique*, portant sur la nature de la signification. Corrélativement l'antiréalisme est une thèse sémantique, qui affirme que la nature de la signification n'est pas indépendante de la connaissance que nous en avons. Il ne s'agit donc pas seulement de dire que nos conceptions de la signification ou de la forme d'une théorie de la signification ont des « implications » métaphysiques réalistes ou antiréalistes, mais aussi que ces conceptions sont partie intégrante de nos thèses métaphysiques elles-mêmes. C'est en ce sens notamment que la « philosophie du langage » est la philosophie première.

L'argument (a)-(c) est encore peu clair sous sa forme schématique, tant que nous n'avons pas précisé en quoi consiste exactement l'alternative entre le réalisme et l'antiréalisme, et en quoi cette alternative se rattache à l'argument (1)-(3) portant sur la compréhension du langage. Commençons par essayer de rendre plus explicite ce dernier argument. La prémisse de base de cet argument est celle que Dummett partage avec Davidson :

1. Cf. aussi Tennant, 1987 : 20 ; Wright, 1976 ; Apphia, 1986.

(i) une théorie de la signification est une théorie de la compréhension : elle doit nous dire ce que c'est, pour un locuteur d'un langage, que comprendre ce langage[1].

Une autre prémisse, implicite, que l'antiréaliste accorde provisoirement à son adversaire, est que

(ii) la connaissance de la signification d'une phrase consiste dans la connaissance de ses conditions de vérité : savoir ce que signifie une phrase c'est savoir quelles sortes d'états de choses sont ou seraient réalisés si cette phrase est (était) vraie.

La prémisse I / impose un réquisit sur la nature de la compréhension de la signification, que nous pouvons appeler le principe de la publicité de la signification. Il est lui-même la conjonction de deux principes[2] :

(A) le principe d'*acquisition* : nous ne pouvons pas former la compréhension d'une phrase, ni saisir ses conditions de vérité si ces conditions de vérité doivent être conçues comme indépendantes de notre expérience, acquises indépendamment de notre observation du comportement humain, non confirmées par ce comportement.

(M) le principe de *manifestabilité* : la signification doit être rendue manifeste par le comportement observable ; elle doit consister dans des conditions reconnaissables de manière publique.

La véritable nature de ces conditions ne pourra être clarifiée que si l'on élucide ce qu'est le « comportement observable » en question. On laissera ce point en suspens. Ces réquisits forment ce que l'on peut appeler la connaissance de l'usage des phrases. A ce stade, l'antiréaliste n'affirme pas que cette connaissance de l'usage est incompatible avec la connaissance des conditions de vérité, mais que cette dernière est contrainte par les conditions d'acquisition et de manifestation qui pèsent sur l'usage : on ne peut pas saisir les conditions de vérité si elles ne sont pas associées à la connaissance de l'usage des phrases. Cela veut dire que la seule connaissance des conditions de vérité ne suffit pas à assurer une connaissance de la signification : quelqu'un peut connaître ces dernières sans être capable d'utiliser les phrases correspondantes, c'est-à-dire sans être capable de manifester sa connaissance par un comportement observable quelconque. Mais comment la signification peut-elle être ainsi publique, acquise et manifeste ? Une autre prémisse de base de Dummett est :

1. Dummett, 1973 : chapitre 1 ; 1975 : 100-101 ; 1976 : 69 ; 1990, chapitre 4.
2. Cf. principalement Dumett, 1976. Je m'appuie ici sur Wright, 1987 : 12-23 ; Bilgrami, 1986 ; Tennant, 1987, chapitre 1, Appiah, 1986.

(I) La connaissance d'un langage est une connaissance *implicite*, et non pas explicite, ou propositionnelle, des significations (conditions de vérité).

Dummett n'a rien à objecter contre l'idée que la compétence sémantique d'un locuteur puisse être représentée sous la forme d'un ensemble de propositions établissant de manière explicite ce que connaît un locuteur. Mais cette représentation n'est pas identique à la connaissance en question ; elle est la « représentation théorique d'une capacité pratique » que les locuteurs ont de reconnaître la signification (les conditions de vérité) des phrases[1]. Pourquoi ne peut-elle être une connaissance explicite ? D'une part parce que du fait que l'on connaisse un langage il ne s'ensuit pas qu'on soit en mesure de formuler explicitement de manière verbale ce en quoi consiste cette connaissance, et d'autre part parce que le but d'une représentation théorique est d'expliquer ce que celui qui ne comprend pas encore un langage doit acquérir pour le connaître : il serait donc circulaire de soutenir que celui qui veut apprendre une langue doit avoir *déjà* une telle représentation théorique[2]. Mais cela ne nous dit pas en quoi consiste cette connaissance implicite :

(R) la connaissance implicite d'un langage consiste dans une capacité pratique à reconnaître les significations : l'aptitude à reconnaître les circonstances qui satisfont, ou ne satisfont pas aux conditions de vérité[3].

Il serait contradictoire avec les réquisits de publicité (A) et (M) que cette capacité pratique de recognition ne puisse pas être acquise et manifestée : on doit donc exiger qu'elle soit rendue manifeste par une autre capacité, par laquelle le locuteur est en mesure de *décider*, par une procédure finie, quelles sont les conditions de vérité d'une phrase[4]. La nature de cette

1. Dummett, 1975 : 100 ; 1976 : 70-71 ; 1985 ; 1986 ; 1990 : 95-96.
2. Dummett, 1976 : 70, 80.
3. Dummett, 1976 : 71 ; 1985.
4. Dummett, 1976 : 81.

capacité ne pourra être précisée que si nous savons exactement en quoi peut consister cette procédure de décision. Mais il est clair d'ores et déjà que cette condition impose un réquisit très fort sur la nature de la capacité pratique de recognition en question :

(C) La capacité pratique de recognition de la signification doit consister en une procédure manifestable de décision des conditions de vérité.

La procédure de décision doit donc être manifeste. Dummett ne requiert pas seulement que le sens soit manifeste dans l'usage ; il requiert aussi qu'il soit *intégralement* manifeste dans l'usage. La connaissance implicite du langage doit être exhaustivement manifestée par l'exercice des capacités pratiques en quoi elle consiste.

Bien que nous puissions, à ce stade de l'argumentation, souhaiter des éclaircissements, le raisonnement n'est supposé faire appel qu'à des principes très plausibles, susceptibles de découler des deux principes (i) et (ii) admis par le tenant d'une conception véri conditionnelle. Si ce dernier peut contester la description qui est faite des réquisits (A), (M), (I), (R) et (C), peut-il sérieusement contester que comprendre un langage soit une capacité pratique, susceptible d'être apprise et manifestable ? Ce sont ces évidences mêmes qui formaient le noyau des conditions de Davidson (§ 1.2).

Il y a pourtant un conflit entre la conception véri conditionnelle représentée par (i) et (ii) et les autres principes. Car s'il est tout à fait naturel d'admettre que connaître les conditions de vérité d'une phrase, c'est comprendre sa signification, la proposition converse est loin d'être évidente. Car si quelqu'un comprend une phrase, il doit posséder une capacité recognitionnelle à saisir les conditions dans lesquelles cette phrase serait vraie. Or la conception véri conditionnelle n'implique rien de tel : elle présuppose non seulement que les phrases ont des conditions de vérité *indépendamment* de notre capacité à reconnaître ces conditions quand elles sont réalisées, mais aussi ((2) ci-dessus) qu'il y a un grand nombre de phrases que nous comprenons, bien que nous n'ayons aucun moyen de savoir dans quelles conditions elles sont vraies ou fausses.

Les exemples de ce genre de phrases, dans le langage naturel, abondent : les phrases contenant des quantifications sur des domaines infinis, les phrases portant sur le futur ou le passé, ou sur des régions inaccessi-

bles de l'espace et du temps, les phrases conditionnelles contrefactuelles, les phrases contenant des termes dispositionnels, ou encore les phrases attribuant à autrui des états mentaux[1]. Comment, par exemple, pouvons-nous déterminer les conditions de vérité de phrases comme « Dupond était brave », ou « Louis XVI avait 8 de tension le 21 janvier 1793 » ? Dummett appelle ces phrases « non effectivement décidables »[2]. On peut dire simplement que ce sont des phrases qui excèdent ou transcendent nos capacités de recognition de leurs conditions de vérité, ou des phrases *transcendantes*.

Il est tentant, à ce point, de raisonner ainsi : puisque la connaissance de la signification consiste dans la connaissance des conditions de vérité des phrases (par (i)), et puisqu'il y a des phrases dont nous ne pouvons pas connaître les conditions de vérité, alors ce sont des phrases dont nous ne connaissons pas la signification. Les phrases « effectivement décidables » dont parle Dummett semblent bien être des phrases vérifiables, et douées de signification parce que vérifiables (en pratique ou en principe). Le raisonnement antiréaliste serait alors simplement une version du raisonnement traditionnel du positivisme vérificationniste viennois, selon lequel il faut distinguer les phrases susceptibles d'être vérifiées ou falsifiées, d'une part (donc pourvues de sens), des phrases qui ne peuvent être ni vérifiées ni falsifiées (donc dénuées de sens). Mais si l'antiréalisme est bien, comme on le verra, une forme de vérificationnisme, il n'est pas de l'espèce positiviste en ce qui concerne sa conception de la signification. Dummett ne nie pas que nous comprenions les phrases transcendantes, et il est d'accord sur ce point avec le partisan de la conception véri conditionnelle. Il est parfaitement compatible avec sa position de dire, par exemple, que nous comprenons la signification de la conjecture de Golbach ou de l'énoncé du théorème de Fermat, bien que nous n'ayons aucun moyen (peut-être seulement présentement) de déterminer les conditions dans lesquelles ces énoncés seraient vrais. Il ne s'agit pas plus de dire que « Dupond était brave » ou « Louis XVI avait 8 de tension le 21 janvier 1793 » sont des phrases dont nous ne pouvons pas comprendre réellement la signification.

1. Dummett, 1976 : 86, 89-101 ; 1978 : 215-247 et 358-374.
2. Dummett, 1976 : 80.

Certains commentateurs de Dummett, comme Mc Ginn, ont été conduits à lui attribuer cette forme de vérificationnisme, parce qu'ils tendent à confondre le fait qu'une phrase ait une *valeur* de vérité décidable, avec le fait que cette phrase ait une signification ou des *conditions* de vérité décidables, et c'est sans doute la thèse (C), qui semble impliquer que la connaissance de la signification consiste dans la capacité à décider de la vérité d'une phrase, qui encourage cette confusion. Mais, comme le dit Tennant, (C) est la manifestation de la compréhension, elle n'est pas ce en quoi celle-ci consiste. En d'autres termes, la saisie de la signification ne garantit pas une saisie de la valeur de vérité[1].

Le raisonnement antiréaliste est donc bien conforme à la prémisse 2 /. Mais alors le tenant de la conception vériconditionnelle est confronté à un dilemme : ou bien il maintient la connexion présumée entre la connaissance de la signification et la connaissance des conditions de vérité, mais doit admettre que nous avons une capacité à connaître les conditions de vérité qui n'est pas manifestable par une capacité de recognition quelconque (sauf à nous attribuer des pouvoirs surhumains), ou bien il renonce à affirmer cette connexion, mais doit admettre que la connaissance de la signification d'une phrase ne consiste pas essentiellement dans la connaissance de ses conditions de vérité. Dans ce dernier cas, il doit rejeter l'idée que le concept de vérité joue un rôle central dans la connaissance de la signification, et doit donc abandonner le présupposé principal de ce que Dummett appelle le réalisme en théorie de la signification. C'est la voie suivie par l'antiréaliste quand il choisit la conclusion 3 / de l'argument : la connaissance de la signification d'une phrase ne peut pas consister dans la connaissance de conditions de vérité non manifestables. Nous ne sommes pas en droit de conclure immédiatement qu'elle consiste dans la connaissance d'autre chose que des conditions de vérité, car la connaissance de certaines conditions de vérité est manifestable : celles des phrases « décidables » ou non transcendantes portant sur les objets usuels qui nous entourent, sur des domaines d'objets mathématiques décidables, ou peut-être sur nos propres états mentaux ou sensations[2]. Dans

1. Mc Ginn, 1980 : 29 ; Tennant, 1987 : 113-115.
2. Dummett, 1976 : 95.

ces cas, il n'y a pas de différence de principe entre la conception vériconditionnelle de la signification, et la conception antiréaliste. Mais il semble bien que l'idée que la connaissance des conditions de vérité doive être en principe manifestable soit étrangère à la conception vériconditionnelle elle-même, parce que celle-ci est indifférente à la distinction entre les phrases transcendantes et les autres. Dans le cas des premières, le vériconditionnaliste paraît incapable d'expliquer en quoi consiste la connaissance des conditions de vérité. Comme le dit Wright :

> En général nous devons distinguer au moins quatre catégories de phrases auxquelles nous devrions ordinairement appliquer la notion de vérité : celles qui sont effectivement décidables ; celles dont nous concevons les conditions de vérité de telle manière que, si elles sont vraies, elles doivent (au moins en principe) être reconnues comme vraies, bien que la connexion correspondante entre fausseté et falsifiabilité soit absente ; celles dont nous ne sommes pas en position d'exclure la possibilité de leur vérification (le théorème de Fermat, ou par exemple toute attribution d'une propriété dispositionnelle dont nous pouvons pas effectivement déterminer les circonstances de révélation mais que nous pouvons reconnaître si elles surviennent) ; et celles dont la vérification excède toujours nos pouvoirs. Pour un réaliste, ces différences ont peu de conséquences pour la notion de compréhension ; connaître la signification d'une phrase de l'une ou l'autre de ces espèces c'est savoir qu'elle est vraie sous certaines circonstances spécifiées, et fausses d'autres. Mais prenez n'importe quel exemple dans la quatrième catégorie : qu'est-ce qui dans un tel cas est supposé constituer cette connaissance alléguée ? Le seul candidat apparent est : une capacité à formuler une analyse conventionnellement correcte de ces circonstances. Mais cette capacité... est indépendante de l'aptitude que peut avoir quelqu'un à utiliser la phrase correctement en réponse à son expérience. Si cette dernière aptitude est essentielle à une compréhension de la phrase, le fait demeure qu'il ne peut y avoir aucune expérience de la vérité de la phrase. Si la compréhension linguistique, par conséquent, est conçue comme étant essentiellement une aptitude pratique, nous n'avons pas d'autre choix que de conclure que dans ces cas il n'y a rien de tel qu'une connaissance des conditions de vérité ou bien que celle-ci réside dans une aptitude qui n'est ni nécessaire ni suffisante pour saisir le sens de la phrase en question. Dans un cas comme dans l'autre, nous abandonnons l'idée que c'est en termes de la connaissance des conditions de vérité que la théorie de la signification doit dépeindre notre compréhension des phrases déclaratives (Wright, 1987 : 55).

Cela nous mène à 3 /, la conclusion de l'argument antiréaliste. Quelle autre candidat que les conditions de vérité peut jouer le rôle de ce qui

est connu quand on connaît la signification ? Seul un trait se rattachant à notre usage des phrases peut faire l'affaire. Dummett propose que ce trait soit notre capacité à vérifier les phrases, et que l'idée réaliste d'une connaissance des conditions de vérité soit remplacée par celle d'une connaissance des conditions de vérification :

> Ce que nous apprenons quand nous apprenons à utiliser ces phrases n'est pas ce que c'est pour elles que d'être vraies ou fausses, mais plutôt ce qui compte pour nous comme établissant de manière concluante qu'elles sont vraies ou fausses : les notions centrales d'une théorie de la signification doivent, par conséquent, être celles de vérification et de falsification plutôt que celles de vérité et de fausseté (Dummett, 1973 : 467).

Quelle serait alors, selon cette conception, la nature de notre connaissance de la signification ? Là où les conditions de vérité sont effectivement décidables, elles coïncident avec les conditions de vérification. Là où elles ne le sont pas, les conditions de vérité ne peuvent pas être ce que nous comprenons. Seules les conditions de vérification jouent ce rôle.

Cet argument suscite une réaction courante de la part du tenant de la conception véri conditionnelle. Celui-ci peut admettre l'incapacité où nous sommes de vérifier effectivement les phrases transcendantes, mais nier qu'elle implique que nous connaissions autre chose que leurs conditions de vérité. Il peut soutenir que ce que nous connaissons, dans ces cas, sont des conditions de vérification idéales, qui pourraient être connues par un être dont les pouvoirs excéderaient énormément les nôtres, et dont nous concevons les conditions de vérification réelles par analogie avec ces conditions idéales. Dans la mesure où elles seraient idéales, ces conditions de vérification seraient identiques aux conditions de vérité[1]. Cette conception, selon Dummett, ne romprait pas avec le réalisme, tout en reposant sur l'acceptation implicite de l'idée qu'il y a un lien étroit entre la signification d'une phrase et sa vérification. Mais elle rencontrerait le même problème que la thèse initiale : car ces conditions idéales de vérification ne pourraient pas plus être manifestables que les conditions de vérité indétectables du réaliste. La seule possibilité est donc de s'en

1. Dummett, 1973 : 466 ; 1986 : 152.

tenir à l'idée que la vérification doit être le produit de nos capacités effectives. En quoi consiste alors la connaissance de la signification des phrases transcendantes ? En rien d'autre que le fait qu'elles aient, ou non, certaines conditions de vérification :

> Une phrase indécidable est simplement une phrase dont le sens est tel que, bien que dans certaines situations reconnaissables nous la reconnaissions comme vraie, dans d'autres comme fausse, et où cependant dans d'autres aucune décision ne soit possible, nous ne possédons aucun moyen effectif pour déterminer une situation qui soit de l'une ou l'autre espèce. Plutôt que de faire appel à des facultés hypothétiques, que nous ne possédons pas, conçues par analogie avec celles que nous possédons, qui nous permettraient de convertir ce genre de phrase indécidable en phrase décidable, nous devrions décrire les choses telles qu'elles sont. Le fait réel de notre pratique linguistique est que les seules notions de vérité et de fausseté que nous avons pour ce type de phrase sont celles qui ne nous autorisent pas à considérer la phrase comme vraie ou fausse de manière déterminée indépendamment de notre connaissance (Dummett, 1973 : 467-468).

L'antiréalisme de Dummett est donc bien une forme de vérificationnisme. Mais ce vérificationnisme n'est pas une prémisse de l'argument dirigé contre la conception véri conditionnelle ; il en est une conséquence, que l'on doit tirer de l'inadéquation de la conception véri conditionnelle.

Cela complète la présentation initiale que l'on peut donner de l'argument négatif que Dummett dirige contre la théorie véri conditionnelle de la signification. Mais comme on l'a vu, sa cible est en fait beaucoup plus large : d'une part, la *reductio* dummettienne procède par assimilation de cette conception véri conditionnelle au réalisme en général, et non pas seulement au réalisme quant aux significations, et d'autre part elle fait partie d'un programme positif d'analyse de la signification dont cet argument négatif n'est qu'une première étape. Il y aurait beaucoup à dire sur la logique propre à l'argument antiréaliste standard, mais je m'en tiendrai dans ce qui suit à la présentation schématique qui a été donnée ci-dessus[1]. Pour évaluer cet argument général, nous avons donc d'abord besoin de savoir ce qu'est, selon Dummett, le réalisme.

1. Pour des analyses plus précises, cf. Appiah, 1986 ; Tennant, 1987, et les divers essais dans Taylor, 1987.

4.2. Qu'est-ce que le « réalisme » ?

Tel qu'il a été caractérisé jusqu'ici, le réalisme est la thèse, associée par Dummett à toute théorie vériconditionnelle de la signification, selon laquelle la signification d'une phrase consiste dans ses conditions de vérité, et selon laquelle la connaissance de la signification consiste dans la connaissance des conditions de vérité. Le réalisme est donc avant tout une thèse sémantique, et une thèse qui s'applique à une théorie sémantique. Mais Dummett soutient aussi que le réalisme est une thèse métaphysique, portant sur la nature de la réalité et de la connaissance que nous en avons, et qu'il y a un lien étroit entre la thèse sémantique du réalisme et la thèse métaphysique. Le problème est de caractériser adéquatement ce lien. Dummett adopte tour à tour plusieurs critères, visant tous à rattacher réalisme sémantique et réalisme métaphysique. J'en distinguerai, dans cette section, deux qui sont plus spécifiquement sémantiques, et deux autres qui sont plus spécifiquement métaphysiques ou épistémologiques, bien que cette distinction ne soit que de raison, puisque le but de Dummett est de la rendre douteuse.

Selon le critère envisagé jusqu'ici, toute théorie sémantique fondée sur la notion de condition de vérité est réaliste. C'est vrai en particulier des théories-T mises en avant par Davidson. Mais ce doit être vrai aussi des théories sémantiques non homophoniques fondées sur la théorie des modèles et la notion de monde possible. Ces théories acceptent l'idée fondamentale que le concept de vérité puisse être expliqué en termes du schéma vériconditionnel « S est vrai ssi *p* » que Dummett appelle « la thèse d'équivalence »[1]. Le premier critère de base du réalisme d'une théorie sémantique est donc :

(I) (i) que le concept central de cette théorie sémantique soit le concept de vérité
(ii) que le concept soit expliqué en termes du schéma vériconditionnel (T) « S est vrai ssi p »
(iii) que notre connaissance de la signification d'une phrase S consiste dans la connaissance de ses conditions de vérité sous la forme du schéma (T).

1. Dummett, 1959 ; 1973 : 445 ; 1978 : xx-xxii ; 1982 : 435-436.

Ce critère est en quelque sorte interne à la formulation d'une théorie sémantique. Il ne spécifie en lui-même aucune implication métaphysique quant au concept de vérité. Mais cette implication apparaît à partir du moment où l'on reconnaît que (I) nous engage à dire que

(II) Les conditions de vérité d'une phrase sont indépendantes de la connaissance que nous pouvons en avoir, ou par rapport à notre vérification.

C'est ce qui ressort de l'argument 1 /- 3 / de la section précédente. Le second critère fondamental du réalisme est donc l'acceptation de la thèse selon laquelle la vérité et les conditions de vérité d'une phrase sont transcendantes par rapport à notre pouvoir de vérification.

Dans les écrits de Dummett, ces deux critères sont la plupart du temps associés à un autre, la Bivalence. Si, selon (I), la vérité est le concept sémantique fondamental, alors il semble bien que cela entraîne que la valeur sémantique d'une phrase ne peut être que le vrai ou le faux ; et si, selon (II), la vérité ou la fausseté sont indépendantes de notre pouvoir de recognition, il s'ensuit que :

(III) (Principe de Bivalence) Toute phrase (transcendante ou non effectivement décidable) est vraie ou fausse de manière déterminée, indépendamment des moyens que nous pouvons avoir de le reconnaître.

Le principe de Bivalence est significatif du réalisme quand il s'applique aux phrases transcendantes ou non décidables, puisque, comme on l'a vu, pour les phrases décidables, l'antiréaliste n'a aucune raison de nier qu'une phrase soit vraie ou fausse de manière déterminée[1]. A de nombreuses reprises, Dummett soutient que le principe de Bivalence est constitutif du réalisme[2]. Ensuite il admet que si le principe est suffisant pour qualifier une position comme réaliste, il n'est pas nécessaire. Par exemple la mise en doute du principe pour les énoncés présentant des « lacunes de valeur de vérité » (comme ceux contenant des termes singuliers vides) ne fait pas obstacle à une conception réaliste. Ou encore un « neutraliste » au sujet des futurs contingents qui a la même conception que ci-dessus

1. L'occurrence du mot « déterminée » est ici essentielle, parce que nous devons distinguer Bivalence, Tiers Exclu et Principe d'Equivalence. Cf. Dummett, 1978, *Préface*.
2. Dummett, 1973 : 466 ; 1976 : 101 ; 1978 : xxix-xxxii, 175.

concernant le futur, mais qui assimile « vrai » à « correctement assertable » peut, en ce sens très précis, souscrire à la Bivalence au sens (III), sans être réaliste quant au futur[1]. Dummett est donc prêt à dire que le fait d'admettre la Bivalence n'entraîne pas par soi-même le réalisme au sens (B) de la transcendance de la vérité par rapport à la vérification. Mais il est, en tout état de cause, décidé à maintenir l'implication converse :

> Il est vrai que la Bivalence est souvent supposée de manière irréfléchie ; mais quand le réaliste y réfléchit, il y a peu de chances qu'il soutienne qu'il dérive sa conception de la vérité de sa croyance en ce principe. C'est plutôt qu'il se suppose lui-même en possession d'une saisie de ce que c'est pour un énoncé que d'être vrai indépendamment de nos moyens de le reconnaître comme vrai, ou de la question de savoir si nous en avons, et sur la base de cette conception il accepte la Bivalence (Dummett, 1987 : 230).

C'est donc le réalisme au sens (II) qui, de ce point vue, constitue l'engagement fondamental. Quoi qu'il en soit, le critère de la Bivalence joue un rôle essentiel dans le raisonnement antiréaliste, parce que la Bivalence est un principe logique, étroitement lié à l'adoption de la logique classique. Si le lien est aussi étroit que Dummett le soutient, alors le réalisme sémantique au sens de (I) et de (II) implique un réalisme logique, d'après lequel la logique classique est la seule logique correcte, et ce réalisme logique est, au moins partiellement, constitutif du réalisme. Rejeter la Bivalence (ou plus exactement refuser de l'asserter et de la nier[2]), c'est prendre le chemin d'un antiréalisme logique pour lequel la logique classique cesse d'être valable, et pour lequel, selon Dummett, la logique intuitionniste doit être appropriée.

Le quatrième critère du réalisme que l'on trouve chez Dummett est explicitement métaphysique :

> (IV) (Vérité-Correspondance) Si une phrase est vraie, il doit y avoir quelque chose dans le monde en vertu de quoi elle est vraie.

1. Dummett, 1978, *Préface*.

2. L'intuitionniste ou l'antiréaliste ne *rejette* pas le principe de Bivalence, mais refuse de l'asserter *et* de le nier. Car les phrases indécidables ne lui fournissent pas de raisons de le rejeter, car il ne s'estime pas en droit d'inférer la fausseté de l'absence de raisons concluantes. Mais pour plus de commodité je m'exprimerai quelquefois dans ce qui suit de façon lâche en parlant du « rejet » intuitionniste du principe.

Dummett soutient que ce principe n'est que partiellement constitutif du concept de vérité, et qu'il n'est applicable qu'indirectement à ce concept. Son statut est plutôt régulateur :

> C'est-à-dire que nous ne déterminons pas d'abord ce qu'il y a dans le monde, puis décidons, sur la base de cela, ce qui est requis pour rendre chaque énoncé vrai, mais plutôt que ayant préalablement décidé quelle devait être la notion appropriée de vérité pour chaque type d'énoncé, nous concluons de cela à la constitution de la réalité (Dummett, 1976 : 89).

Il y a un domaine où l'application du principe (III) et son lien avec les autres principes sont particulièrement évidents : c'est celui des mathématiques, où la thèse selon laquelle les énoncés mathématiques ont des conditions de vérité indépendantes de notre connaissance conduit à admettre l'existence d'objets mathématiques et une forme de platonisme. Dummett soutient précisément que le platonisme mathématique dérive directement d'une thèse concernant la signification et la vérité des énoncés mathématiques[1].

Qu'il y ait en général, pour toute phrase vraie, quelque chose en vertu de quoi elle est vraie est la thèse que Dummett qualifie parfois de « réalisme naïf ». Mais toute forme de réalisme n'est pas nécessairement naïve. Certaines formes de la thèse peuvent au contraire être associées à l'inexistence des objets correspondant à une classe considérée, lorsque l'on a affaire à une forme de réductionnisme. Par exemple le phénoménisme quant aux objets matériels réduit ceux-ci à des classes de sensations, et semble donc être une forme d'antiréalisme. Mais les énoncés de la classe réductrice ont des conditions de vérité bivalentes. De même le matérialisme identitaire qui réduit les états mentaux à des états physiques. Cette forme de réalisme réductionniste est ce que Dummett (1982a) appelle un réalisme « sophistiqué ».

Chacun de ces critères permet de qualifier une position comme « réaliste ». Mais Dummett ne soutient pas que le réalisme soit, de prime abord, une doctrine unifiée. On n'est pas en général réaliste tout court, mais relativement à telle ou telle classe de phrases, et, par là même, relative-

1. Cf. en particulier Dummett, 1973.

ment à certaines classes d'entités. Ainsi en mathématiques relativement aux objets mathématiques, en éthique relativement aux valeurs, ou encore en métaphysique relativement à l'existence du passé ou du futur, ou relativement à l'existence des objets matériels. Dummet soutient que chacun de ces réalismes locaux rencontre des problèmes spécifiques, mais il n'implique pas que si l'on est réaliste dans l'un ou l'autre de ces secteurs, on doit l'être dans d'autres, et parallèlement en ce qui concerne la thèse opposée antiréaliste. Il admet néammoins qu'il existe, pour chaque réalisme local qui prétend être en possession d'une certaine conception des conditions de vérité des énoncés appartenant à une classe donnée, un défi, que l'antiréaliste adresse au réaliste, consistant à demander en quoi consiste réellement cette conception. L'antiréalisme global est alors la croyance générale que ce défi ne peut être vraiment relevé que pour des énoncés pour lesquels nous possédons une méthode effective pour déterminer s'ils sont vrais ou faux[1].

L'antiréalisme dummettien n'est pas la seule forme d'antiréalisme contemporain fondé sur des considérations sémantiques. Il existe deux autres types de conceptions. La première admet que, pour une région de discours particulière, le réalisme sémantique au sens de Dummett est correct, et que les énoncés de cette région ont des conditions de vérité. Mais selon ce type de théoricien, la classe d'énoncés en question est simplement fausse : les entités qu'ils désignent n'existent pas, et c'est simplement une *erreur* que de croire en leur existence. Ce genre de « théorie de l'erreur » a été défendue pour les énoncés moraux par Mackie (1977), et pour les énoncés mathématiques par Field (1981, 1989). Selon une conception, les énoncés d'une région donnée ne sont pas en fait vrais ou faux, mais *semblent* seulement l'être : ils *expriment* des attitudes ou des sentiments. Ce paradigme *expressiviste*, dont l'expression contemporaine remonte à Ayer (1936) et aux positivistes, a été défendu plus récemment en éthique notamment par Blackburn (1985, 1993). Ces formes d'antiréalisme présentent certaines affinités avec la position instrumentaliste en philosophie des sciences, selon laquelle les termes théoriques ne désignent pas des entités réelles, mais servent d'instruments pour des prédictions[2].

1. Dummett, 1982 : 432 ; 1982a : 55.
2. Cf. en particulier Van Fraassen, 1980, pour une expression de cette conception qui repose en partie sur des considérations « pragmatiques ».

L'antiréalisme n'est donc, pas plus que le réalisme, une doctrine unifiée. Mais ici, quand nous ferons référence à ces doctrines, nous les comprendrons dans leur sens dummettien.

4.3. Holisme et molécularisme

Dummett dissocie rarement sa critique du réalisme de la critique d'une autre thèse portant sur la signification, le holisme. Mais il ne donne en général aucune explication systématique de cette doctrine. Comme au § 1.2, nous pouvons distinguer le holisme de la phrase (la signification d'un mot ou d'une expression est fonction du rôle qu'il joue au sein d'une phrase) du holisme du langage (la signification d'une phrase d'un langage dépend toujours de la signification des autres phrases de ce langage). Dummett n'a aucune objection contre le holisme de la phrase, qui est une version du principe fregéen de contextualité, qu'il considère comme un principe fondamental de toute théorie sémantique[1]. Le holisme du langage a, selon lui, au moins deux versions distinctes. La première est ce que l'on peut appeler le holisme *épistémologique* ou de l'« inextricabilité », dont l'expression la plus caractéristique se trouve chez Quine : c'est la thèse selon laquelle le rejet de la distinction analytique/synthétique nous conduit au rejet de toute distinction entre des phrases théoriques et des phrases observationnelles, et à l'idée que « la signification de chaque phrase est contaminée par la théorie », et par conséquent par celle de toutes les autres phrases[2]. La seconde version est ce qu'il appelle le holisme *constitutif* : c'est l'idée que la signification d'une expression isolée n'existe que relativement à un réseau complexe et est constituée par sa place dans ce réseau, en sorte qu'il n'est pas possible de donner un contenu défini à une phrase individuelle, sinon par l'intermédiaire du langage tout entier[3]. Il semble bien que ce soit le holisme constitutif que Dummett vise également quand il le caractérise au moyen non plus de la métaphore du réseau, mais de celle du jeu, et le qualifie ainsi de « radical » :

1. Dummett, 1982, chapitre 19. Tennant, 1987, chapitre 5.
2. Dummett, 1978 : 134 ; 1990 : 242, sq. Je m'appuie sur Tennant, 1987, chapitres 5-6.
3. Dummett, 1978 : 48.

> Un jeu de stratégie donne un excellent modèle pour une théorie radicalement holiste du langage. Celui qui apprend un jeu doit absolument apprendre les règles une à une ; mais il ne peut pas saisir la signification d'un coup particulier dans le jeu sans recourir à toutes les règles. Or une position ou un coup dans le jeu a sans aucun doute une signification : et sa signification dépend systématiquement de toutes les places qu'occupent les pièces du jeu et de leurs pouvoirs. En ce sens, donc, l'analogue du principe selon lequel la signification d'une phrase est déterminée par sa composition est satisfait ; et il n'entre nullement en conflit avec le caractère holistique de la signification d'un coup particulier. La raison est celle sur laquelle Frege insiste abondamment..., à savoir que, bien qu'un coup dans un jeu ait une signification — la signification que lui donnent les règles du jeu — il n'a pas de contenu : il n'exprime pas, dans la terminologie de Frege, de pensée. La seule analyse que l'on puisse donner de sa signification est par conséquent sa place dans l'arbre de tous les coups possibles ; et de cela, un joueur ne peut pas avoir plus qu'une saisie partielle. Si les joueurs n'avaient jamais aucune saisie de la signification des coups, le jeu ne serait jouable que comme un jeu de hasard, pas de stratégie ; s'ils en avaient une saisie totale, le jeu ne serait pas jouable du tout, comme quand les adultes jouent au morpion (Dummett, 1987 : 247-248, cf. aussi 1978 : 135).

Quel est le lien entre le holisme du langage et le réalisme ? Comme le réaliste, pour lequel la signification d'une phrase est indépendante des conditions de sa vérification, le holiste soutient que la signification d'une phrase ne peut être connue isolément. La vérification des phrases ne peut donc être que globale, et parce qu'elle est globale, elle devient transcendante par rapport à nos moyens effectifs de vérification. De même elle n'est pas manifestable, sinon par l'usage du langage tout entier. L'analogie est donc étroite avec la thèse selon laquelle les conditions de vérité n'ont pas besoin d'être connues pour que les phrases aient une signification. Mais rien de ceci ne montre que le réalisme implique le holisme, ou inversement.

Au holisme, Dummett oppose ce qu'il appelle le *molécularisme*, c'est-à-dire la doctrine selon laquelle (a) chaque phrase a une signification isolée, indépendamment de la signification des autres phrases du langage, et (b) selon laquelle la signification d'une phrase ne dépend pas de celle de phrases plus complexes qu'elle. Selon le molécularisme,

> Chaque phrase possède un contenu individuel qui peut être saisi sans une connaissance du langage tout entier... Chaque phrase retient son contenu, est utilisée exactement de la même manière que nous l'utilisons présentement, même

> si elle appartient à un langage fragmentaire extrêmement rudimentaire, contenant seulement les expressions qui figurent dans cette phrase et dans d'autres, du même niveau ou d'autres niveaux, dont la compréhension est nécessaire pour la compréhension de ces expressions ; dans un tel langage fragmentaire, des phrases de plus grande complexité que celles données ne figureraient pas (Dummett, 1978 : 302-303).

Le molécularisme est parfaitement compatible avec le holisme de la phrase ; il en est même l'expression directe. Tous deux s'opposent à l'*atomisme*, la thèse selon laquelle la signification d'un mot peut être déterminée indépendamment de celles des phrases ou des expressions complexes dans lesquelles il figure[1]. En ce sens, le molécularisme est une autre formulation du principe fregéen de compositionnalité, et de la thèse selon laquelle la phrase est l'unité sémantique de base. Ce que Dummett ajoute à ces principes est l'idée qu'il y a un ordre partiel dans la compréhension des phrases : la signification de phrases atomiques ne peut pas dépendre d'autre chose que ses éléments et pas de celle de phrases plus complexes, et des phrases complexes ne peuvent pas avoir une signification qui dépendrait de celles de phrases plus complexes. Le holisme au contraire refuse l'idée d'une telle asymétrie dans la compréhension[2].

Le contexte dans lequel le conflit entre le holisme et le molécularisme est le plus explicite est celui de l'inférence logique ou déductive, que Dummett expose dans « The Justification of Deduction » (1975). Selon une conception molécullariste de l'inférence logique, on doit pouvoir donner un sens à l'idée d'une justification des règles déductives. Elles sont justifiées « précisément... par le fait qu'elles demeurent fidèles aux contenus individuels des phrases qui surviennent dans toute déduction effectuée en accord avec ces règles » (*ibid.*, 303). On justifie habituellement les règles par un théorème de complétude établissant que les règles d'inférence préservent la vérité et la validité. Mais pour un antiréaliste, pour qui le concept central est celui d'assertabilité et non pas de vérité, ce n'est pas suffisant. On doit exiger que l'introduction d'une nouvelle constante étende le lan-

1. Dummett, 1976 : 72 ; nous avons rencontré cette doctrine au § 2.6 au sujet de Fodor. Il est surprenant que Fodor et Le Pore, 1992, s'appuient si peu sur la critique du holisme par Dummett (qu'ils ne touchent que p. 8-10).

2. Dummett, 1987 : 72.

gage de façon *conservatrice*. En d'autres termes, on n'a pas le droit d'asserter des phrases dans lesquelles cette constante n'apparaît pas, et pour lesquelles on n'avait pas de justification avant l'introduction de cette constante[1]. Une conception holiste de l'inférence logique ne requiert aucune justification de la déduction ; elle considère les règles logiques comme justifiées par l'ensemble d'une pratique linguistique. Transposée à la théorie de la signification en général, le molécularisme exige donc que toute introduction d'un fragment à un langage soit une extension conservatrice de ce langage.

Indépendamment de l'affirmation de la thèse molécullariste, on peut voir en quoi le rejet du holisme est important pour renforcer l'argument antiréaliste standard. Car cet argument appelle précisément une réponse simple, inspirée par le holisme. Pourquoi, peut-on dire, voir une difficulté particulière dans le fait que certaines phrases de notre langage ont des conditions de vérité transcendant nos pouvoirs de vérification ? Nous comprenons ces phrases, nous sommes capables de les employer avec succès, et cela peut bien suffire à manifester notre compréhension. De plus le phénomène de la compréhension de phrases que nous ne sommes pas en mesure de vérifier n'est pas isolé. Il se produit constamment, notamment sous la forme de ce que Putnam (1975) appelle « la division du travail linguistique » : nous déférons à des experts la fixation de la signification exacte de termes dont nous serions, si on nous le demandait, bien incapables d'expliquer la signification. Pourquoi exiger que nous devions reconnaître toutes les conditions de vérité de nos phrases ? En parlant notre langage, nous nous engageons dans une pratique qui justifie par elle-même le fait que nous attribuions des conditions de vérité indépendantes à nos phrases[2]. Le molécularisme est précisément destiné à rejeter cette réponse facile à l'argument antiréaliste. Il consiste à demander de quel droit le holisme peut se dispenser d'une explication précise de la signification, et d'une justification de notre pratique linguistique.

Il importe cependant de bien voir la différence entre l'argument antiholiste et l'argument antiréaliste standard examiné ci-dessus. Ce dernier

1. Cf. également Dummett, 1990, chapitres 8-10. Je n'examine pas ici ces conceptions, que j'ai discutées dans Engel, 1989, chapitre XII et 1991c.

2. Dummett, 1978 : 427-428 ; cf. Bilgrami, 1986 : 104-105.

associe la transcendance des conditions de vérité par rapport aux conditions de vérification ou d'usage à la présence, dans le langage, de phrases transcendantes ou indécidables dont le réaliste est incapable de représenter les conditions de reconnaissance. Dans l'argument antiholiste au contraire, il n'est pas nécessaire de supposer que *certaines* phrases dépassent nos capacités de recognition : d'après la conception holiste, *toute* condition de vérité d'une phrase du langage est vouée, en un certain sens, à échapper à nos capacités de recognition, puisque sa signification ne peut être attestée que relativement aux conditions de vérité de toutes les autres phrases. En ce sens, le holisme apparaît comme une doctrine beaucoup plus radicale que le réalisme. Dummett va jusqu'à dire que le holisme n'est même pas « une théorie de la signification » ; il en nie la possibilité même :

> Dans la conception holiste, aucun modèle du contenu individuel d'une phrase ne peut être donné : nous ne pouvons pas saisir le pouvoir représentatif d'une phrase quelconque sinon par une saisie complète des propensités linguistiques sous-tendant notre usage du langage tout entier ; et, quand nous avons une telle saisie de la totalité, il n'y a aucun moyen par lequel ceci puisse être systématisé de manière à nous donner une vision claire de la contribution d'une partie donnée de l'ensemble. Aucune phrase ne peut être considérée comme disant quoi que ce soit par elle-même : la plus petite unité qui puisse être tenue comme disant quelque chose est la totalité des phrases crues, à un moment donné, être vraies ; et de cette totalité complexe, aucune représentation n'est possible — nous faisons partie du mécanisme, et ne pouvons l'examiner de l'extérieur (Dummett, 1978 : 309).

4.4. L'antiréalisme dummettien

Nous avons jusqu'ici considéré essentiellement la partie négative de l'argument antiréaliste. Envisageons à présent la doctrine positive. Comme on le verra, celle-ci diffère considérablement de Dummett à ses disciples. Il ne s'agira ici que de la position dummettienne, présentée assez schématiquement, et que je devrai détacher des analyses particulières qui motivent principalement son programme, qui portent notamment sur la réalité du passé (Dummett, 1969) et sur les mathématiques (Dummett, 1973, 1978).

Comme la conception vériconditionnelle à laquelle il s'oppose, l'antiréalisme sémantique suppose que l'on peut construire une théorie systématique de la signification destinée à expliquer la compétence sémantique des locuteurs. Ses réquisits de base seront ceux de manifestabilité et d'acquisition de la signification. Dummett considère, comme Davidson, que la notion de signification gagne à être élucidée non pas en termes d'une analyse directe de la notion de signification, mais en termes d'une analyse de la forme des théories de la signification[1]. Comme la conception vériconditionnelle, une théorie antiréaliste de la signification devra faire appel à un concept central susceptible de servir de valeur sémantique aux expressions du langage.

Ici la réponse de Dummett a varié. Dans un premier temps, dans « Truth » (1959), il soutient que le concept de signification ne peut pas être élucidé au moyen de celui de conditions de vérité, et qu'il faut par conséquent rejeter la notion même de vérité comme notion centrale en sémantique. Il défend, dans cet article, la conception de la vérité-redondance (§ 3.6), selon laquelle il n'y a rien de plus dans la notion de vérité que l'expression de l'équivalence entre « Il est vrai que *p* » et « *p* ». Selon lui, le schéma tarskien « *S* est vrai ssi *p* » n'est pas autre chose que l'expression de cette « thèse d'équivalence ». Mais il est incompatible avec une explication du sens du prédicat « vrai » en termes de conditions de vérité et de fausseté des phrases par leurs tables de vérités classiques usuelles[2]. Dummett en conclut que l'explication de la signification des constantes logiques usuelles en termes de condition de vérité, qu'il assimile à une conception réaliste de la signification pour ces constantes, doit être abandonnée, et remplacée par une analyse en termes des conditions d'assertion :

> Nous apprenons le sens des opérateurs logiques en apprenant à utiliser les énoncés dans lesquels ils figurent, c'est-à-dire en apprenant à asserter ces énoncés sous certaines conditions. Ainsi nous apprenons à asserter « P et Q » quand nous pouvons asserter P et pouvons asserter Q, à asserter « P ou Q » quand nous pouvons asserter P ou pouvons asserter Q... Nous avons ici complètement

1. Cf. la citation de Dummett ci-dessus, § 1.1.
2. Dummett, 1978 : 6.

> renoncé à expliquer la signification d'un énoncé en stipulant ses conditions de vérité. *Nous n'expliquons plus le sens d'un énoncé en stipulant sa valeur de vérité en termes des valeurs de vérité de ses constituants, mais en stipulant à quelles conditions il peut être asserté, et ce en termes des conditions sous lesquels ses constituants eux-mêmes peuvent être assertés* (Dummett, 1978 : 6, tr. fr. 67).

Ceci revient à exposer la signification des constantes logiques en termes intuitionnistes, en rejetant le principe du Tiers Exclu et en refusant d'asserter le principe de Bivalence, et à chercher en général à transférer à tous les énoncés les explications que les intuitionnistes donnent de la signification en logique et en mathématiques. Dans « Truth », Dummett ne voit pas d'autre moyen de justifier ces explications que de renoncer à utiliser la notion même de vérité, et en l'éliminant, selon la théorie de la vérité-redondance.

Dans un second temps[1], Dummett abandonne la théorie de la vérité-redondance et maintient que la notion de vérité doit garder une place importante dans la formulation d'une théorie de la signification. La signification d'une phrase doit bien consister dans ses conditions de vérité, mais l'application de la notion de vérité doit se limiter aux conditions de l'assertion correcte. Il ne s'agit donc pas d'éliminer la notion de vérité, mais de la modifier, de manière à rétablir son lien avec la notion d'assertion :

> La notion de vérité prend sa source dans la notion primitive du caractère correct d'une assertion ; mais elle ne coïncide pas avec elle. C'est un trait intrinsèque de la notion de vérité que nous puissions effectuer une distinction entre la vérité de ce quelqu'un dit, et les raisons qu'il a de penser que c'est vrai : l'idée qu'une assertion est jugée par des critères de correction ou d'incorrection ne donne pas encore de base pour cette distinction (Dummett, 1976 : 84).

Le problème est donc celui de caractériser d'abord la notion de vérité, puis de caractériser les conditions de l'assertion des phrases, c'est-à-dire de leur usage. C'est pourquoi Dummett (1976) propose l'image globale suivante d'une théorie de la signification pour un langage. Le noyau de la théorie sera constitué par une théorie de la vérité, qui spécifiera induc-

1. Dummett, 1976 ; préface à 1978, postface à « Truth », *ibid.* ; cf. aussi Dummett, 1978 : 459-60, et sur cette évolution Baldwin, 1991.

tivement les conditions de vérité du langage, et que Dummett préfère appeler « théorie de la référence », parce qu'elle consistera essentiellement à attribuer des références aux mots individuels. Mais une telle théorie sera insuffisante comme théorie de la signification. Il n'y a, en soi, rien à objecter à l'idée que la signification d'une phrase est constituée par ses conditions de vérité, et à l'idée que la notion de vérité doive être contrainte par le principe d'équivalence « *S* est vrai ssi *p* ». Mais ce principe n'explique le rôle de la notion de vérité qu'à l'intérieur du langage. Il ne peut être appliqué que si nous supposons que le locuteur a déjà une compréhension préalable de la signification des phrases du langage ne contenant pas le mot « vrai ». C'est pourquoi le principe d'équivalence, et la théorie de la vérité-redondance qui se fonde sur lui, sont inadéquats pour expliquer le sens du mot « vrai » : ils ne s'appliquent qu'à des phrases qui sont déjà individualisées du point de vue de leur contenu, c'est-à-dire dont le locuteur connaît déjà la signification. En d'autres termes, la notion de vérité n'est applicable à une élucidation de la signification que dans la mesure où cette notion est *déjà comprise*[1]. Cette critique jouera un rôle essentiel dans les objections que Dummett adresse à Davidson. Mais ce qu'il nous importe ici de noter est qu'elle justifie l'adjonction à une théorie de la signification d'un second étage, supplémentaire par rapport à la théorie de la vérité et de la référence, qui sera constitué par ce que Dummett appelle une théorie du *sens*. Celle-ci sera chargée d'établir en quoi consiste la connaissance, par le locuteur, des conditions de vérité et de référence, en rattachant aux propositions de cette dernière les capacités spécifiques par lesquelles le locuteur manifeste sa connaissance des conditions de vérité. C'est à ce point que la vérité devra se relier à l'usage, et que les conditions de vérité devront être rattachées aux conditions de vérification. Finalement, comme nous l'avons vu au § 3.1, Dummett ajoute à la théorie de la vérité et à celle du sens une théorie de la force des énoncés, et des conditions de l'énonciation des phrases dans des conditions conventionnelles présidant aux divers actes de langage effectuables.

Le fait que, dans l'image tripartite proposée par Dummett, la partie centrale soit constituée par la théorie de la vérité n'implique évidemment

1. Dummett, 1975 ; 1976 : 78 ; 1978 : 460.

pas que le concept de vérité employé soit le concept classique ou réaliste. Dummett propose au contraire d'abandonner ce concept, et le principe de Bivalence qui l'accompagne, au profit d'une explication intuitionniste des constantes logiques, spécifiée en termes non pas de vérité mais en termes de démonstration. Cette explication nous fournit un prototype pour formuler une théorie de la signification dans laquelle les *conditions d'assertion* se seront substituées aux conditions de vérité[1]. Il resterait encore à voir comment ces réquisits généraux s'accordent avec le molécularisme. Comme on l'a vu, cette thèse exige en général que la signification d'une phrase ne soit pas déterminée par celle de phrases plus complexes. Transcrit en termes de conditions d'assertion, cela veut dire qu'une phrase ne peut être assertée que si nous avons les moyens de l'asserter *directement*, sans l'intermédiaire de l'assertion d'autres phrases, ou, si nous l'assertons à partir d'autres phrases (comme une conclusion à partir de prémisses), ces autres phrases ne doivent pas être de complexité plus grande que celle de cette conclusion. Corrélativement, si nous assertons d'autres phrases à partir de cette phrase comme prémisse, ces dernières ne doivent pas être telles que nous puissions asserter quelque chose qui n'était pas, en quelque sorte, contenu dans ces prémisses. Dummett considère qu'on doit ainsi fixer l'usage d'une assertion en fonction d'une part de ses *fondements* ou *conditions*, et d'autre part de ses *conséquences*. Le molécularisme est la thèse selon laquelle il doit y avoir un accord entre les fondements et les conséquences d'une assertion. Ces réquisits sont des transpositions directes des réquisits intuitionnistes d'*harmonie* des règles d'introduction et d'élimination des constantes logiques et d'*extension conservatrice*[2].

Cela ne nous donne encore qu'une idée fort imprécise de ce que peut être une sémantique en termes de conditions d'assertion, mais on laissera là l'exposé du programme de Dummett, pour considérer les objections qu'il adresse à Davidson.

1. Dummett, 1976 : 110-111.

2. Cf. Dummett, 1973 : 355-362 ; 1975 ; 1976 : 117, 1978 : 221 ; l'exposé le plus complet est Dummett, 1990. Cf. Prawitz, 1977, Tennant, 1987, Kremer, 1988, Peacocke, 1987, 1990, Couture, 1991 et les références de la note 25 ci-dessus.

4.5. *Théorie modeste et théorie substantielle*

Bien que les arguments antiréalistes envisagés jusqu'ici s'adressent à toute conception vériconditionnelle de la signification, Dummett considère que Davidson en est le représentant typique, et lui adresse des critiques spécifiques. Ce sont principalement celles qui figurent dans Dummett (1975) que je considérerai ici[1].

La stratégie argumentative de Dummett contre Davidson est inverse de celle que nous avons exposée jusqu'ici : Dummett ne commence pas par assimiler la position de Davidson au réalisme pour ensuite considérer le holisme comme un corollaire de ce réalisme, mais s'attaque directement à son holisme, pour ensuite l'assimiler au réalisme. Il procède en trois étapes. Tout d'abord a / il établit une opposition entre deux sortes de théories de la signification, les théories « modestes », dont celle de Davidson est une espèce — et les théories « substantielles », et soutient que les premières ne peuvent expliquer la signification. Ensuite b / il soutient qu'une théorie modeste implique nécessairement une forme de holisme. Enfin c / il s'avise du fait qu'une théorie de type davidsonien peut en un certain sens être substantielle sans cesser d'être holiste et par conséquent inadéquate et d / précise les raisons de cette inadéquation. Détaillons ces étapes.

a / Dummett qualifie de modeste une théorie qui établit seulement quels concepts un locuteur possède, sans expliquer en quoi consiste la maîtrise de ces concepts, ce qui selon lui limite une telle théorie à simplement systématiser un savoir que le locuteur possède déjà, et présuppose donc que ce savoir ne peut être attribué qu'à quelqu'un qui comprend déjà le langage. Une théorie *substantielle* ou « robuste » (*full-blooded*)[2] vise au contraire à articuler les concepts qui constituent la maîtrise par un locuteur des expressions de sa langue, et à expliquer ces concepts à quelqu'un

1. Dummett est revenu sur ces critiques dans (1987) et (1990), et bien que sur un point important examiné ici il ait reconnu avoir fait une erreur, le fond général de sa critique est resté le même. Sur ces critiques, cf. Laurier, 1991 et Seymour, 1991.

2. Laurier, 1991, préfère dire ici « robuste ». Je garderai « substantiel » pour maintenir l'analogie avec l'opposition substantiel/modeste ou déflationniste dans les théories de la *vérité*.

qui ne les posséderait pas encore. Une TS du type de celles qu'envisage Davidson, prenant la forme d'une théorie-T, apparaît ainsi comme modeste, dans la mesure où, comme il le dit, elle vise à établir quelle connaissance « suffirait » à un locuteur pour comprendre un langage. Comme on l'a vu, selon Davidson, une TS n'est pas un manuel de traduction (§ 1.3). Mais Dummett se demande dans quelle mesure une théorie modeste fait mieux qu'un manuel de traduction, et retourne contre Davidson la critique que fait ce dernier d'une conception traductionnelle de la signification. Car dans l'hypothèse où les phrases-T « donnent la signification » des phrases de *L*, comment un locuteur qui saurait que « "La neige est blanche" est vrai ssi la neige est blanche » pourrait-il *savoir* ce que « la neige est blanche » signifie, s'il ne sait pas quelle *proposition* cette phrase exprime ? La simple connaissance de la vérité d'une phrase-T ne suffit pas ; il faut aussi savoir ce qui *justifie* cette connaissance. Ce point se rattache étroitement à la critique que fait Dummett de la thèse d'équivalence, qui ne peut, selon lui, expliquer la vérité que dans la mesure où ce concept est déjà compris. Une théorie modeste de la signification est donc triviale, et ne peut au mieux que systématiser un savoir déjà maîtrisé. Tout ce raisonnement est sous-tendu par un principe que Dummett explicite ailleurs : une caractérisation de la vérité dans un langage *via* une théorie-T ne peut être qu'« interne » à cette théorie, destinée à quelqu'un qui comprend déjà le langage, alors qu'une explication véritable doit être « externe », c'est-à-dire introduite dans des termes qui ne présupposent pas d'avance une compréhension de la signification. Une théorie est substantielle précisément en ceci qu'elle doit nous permettre d'introduire, et d'apprendre à un novice, les concepts qui constituent la signification[1]. Ailleurs, Dummett soutient qu'une théorie modeste refuse d'attribuer des contenus aux phrases d'un langage « de l'extérieur » parce qu'elle accepte implicitement l'idée que les concepts — et les pensées — qui constituent ces contenus sont indépendants du langage, et par conséquent parce qu'elle souscrit à une conception psychologiste du langage comme un *code* dans lequel un locuteur encode des pensées indépendantes du langage qui les exprime[2]. Une théorie modeste délègue

1. Dummett, 1976 : 77 ; 1980 ; 1987 : 238-239.
2. Dummett, 1985, 1986, 1987.

à une théorie autonome la tâche d'expliquer comment les individus ont des concepts indépendamment de leur langage. Comme dans une théorie traductionnelle, on suppose que les pensées « nues » peuvent être traduites dans un langage, comme si l'on pouvait avoir une conception de ce que c'est que comprendre ces pensées antérieures à leur usage dans un langage. Une théorie modeste se heurte donc à la même objection que celle que Wittgenstein adressait à la conception augustinienne du langage : « Il décrit l'apprentissage du langage humain comme si l'enfant venait dans un pays étrange et n'en comprenait pas la langue ; c'est-à-dire comme s'il avait déjà un langage, mais pas celui-là. »[1]

b / Dummett explique ensuite pourquoi une théorie modeste ne peut être que holiste. Il admet bien qu'en elle-même une théorie-T n'est pas holiste, mais au contraire atomiste et moléculaire, puisqu'elle donne un axiome spécifique à toute expression, et montre comment les expressions complexes sont construites à partir des expressions simples. Mais elle est holiste en ceci qu'elle est incapable d'attacher un contenu spécifique de connaissance à chaque axiome, et n'attribue au locuteur qu'une connaissance des axiomes pris globalement :

> La connaissance qu'a un locuteur des significations d'une phrase individuelle est représentée comme consistant en la saisie d'une partie d'une théorie déductive, et ceci est connecté avec ses énonciations effectives par le fait que la saisie de l'ensemble de la théorie semble entraîner, d'une façon qui n'est pas expliquée, sa commande du langage dans sa totalité ; mais aucune manière n'est donnée, même en principe, de segmenter cette aptitude à utiliser le langage dans son ensemble dans des aptitudes composantes distinctes qui manifesteraient sa compréhension des mots, phrases ou types de phrases individuels. Pour effectuer une telle segmentation, il serait nécessaire de donner une analyse détaillée de l'aptitude pratique en laquelle consiste la compréhension d'un mot ou d'une phrase particulière, tandis que dans la conception holistique non seulement la maîtrise qu'a un locuteur de son langage ne peut pas être ainsi segmentée, mais encore on ne peut pas en donner une description détaillée. Par conséquent, l'articulation de la théorie ne joue aucun rôle véritable dans l'analyse de ce qui constitue la maîtrise qu'a un locuteur de son langage (Dummett, 1975, 116).

Cela revient à dire que la conception modeste d'une théorie de la signification inclut seulement ce que Dummett appelle une théorie de la

1. Wittgenstein, 1951, § 32. Baldwin, 1991, reprend cette critique contre Davidson.

vérité et de la *référence*, et aucune théorie du *sens* capable d'assigner à chaque spécification des conditions de vérité et de référence, une analyse de ce en quoi consiste la *connaissance* de ces conditions de vérité et de référence. De ce point de vue, une théorie modeste est l'équivalent, à une plus large échelle, de la théorie selon laquelle la sémantique des noms propres pourrait se ramener à une simple connaissance de leur référence sans qu'intervienne leur sens ou « mode de présentation »[1].

c / Dummett (1975, appendice) s'avise pourtant de ce que ce diagnostic sur la conception davidsonienne est sans doute incorrect. Il admet que ce qui tient lieu, chez Davidson, de théorie du sens, est constitué par la théorie de l'interprétation, qui a pour rôle de rattacher une TS à la connaissance que peuvent avoir les locuteurs de leur langage, et à la manifestation de cette connaissance. Cela aurait, selon Dummett, pour effet de réduire le holisme de Davidson à un holisme seulement *méthodologique*, d'après lequel les assignations de croyances et de signification de l'interprète radical doivent s'adapter aux données *globales* dont il dispose. Or Dummett ne voit aucune objection à ce holisme, qu'il juge presque « banal », dans la mesure où il ne concerne que les confirmations d'une TS. Mais il soutient que le holisme davidsonien est plus profond, et tient non seulement à la confirmation de la théorie, mais à ce qu'elle décrit. Selon Dummett, le holisme de Davidson a deux composantes. La première est constituée par le fait qu'un locuteur *P* croit, ou connaît, pour toute phrase *S* qu'il tient pour vraie, une phrase-T (T)-*S* correspondante. La seconde composante consiste en ceci que *P* croit ou connaît non pas quelque chose sur *S* ou (T)-*S*, mais *le fait qu'*il y a une théorie-T permettant de dériver (T)-*S* et qui valide le maximum de jugements effectués par *P*. Du point de vue de l'interprète, cette connaissance de la signification de *S* ne peut être attribuée à *P* si et seulement si la théorie-T assignée à son langage réalise la meilleure assignation *totale* (charitable) aux phrases de *L*. C'est une conception, selon Dummett, très similaire à celle de Wittgenstein quand il soutient que le sens du nom propre « Moïse » est constitué par un nombre *suffisant* de jugements vrais au sujet de Moïse[2]. Or cela

1. Dummett, 1975 : 125-126.
2. Wittgenstein, 1951, § 79 ; Dummett, 1975, *130 sq* ; 1987 : 242-243 ; cf. aussi 1973 : 128-129.

revient à dire que l'on ne peut pas avoir de connaissance du sens d'une phrase ou d'un mot quelconque sans avoir une connaissance simultanée de l'*ensemble* des autres jugements sur cette phrase ou ce mot, et donc du langage entier. Il s'ensuit que le holisme de Davidson n'est pas seulement méthodologique, ou propre à la façon dont une théorie-T est testée. C'est un holisme constitutif du réseau, qui rend en fait impossible l'idée qu'un locuteur pourrait avoir la connaissance du sens et de la référence d'une expression individuelle, puisqu'il lui faudrait déterminer toutes les assignations possibles de référence et de vérité — une tâche surhumaine[1]. Même augmentée d'une théorie de l'interprétation, la conception davidsonienne ne cesse pas d'être « modeste » : la modestie ne vient pas tant du fait que l'on tient la connaissance des conditions de vérité des phrases d'un langage comme suffisante pour la connaissance de leurs significations, mais à l'incapacité pour une théorie holistique de déterminer les significations spécifiques que les locuteurs attribuent aux expressions de leur langage.

On pourrait cependant penser que Dummett confond deux points de vue qui sont distingués dans la théorie davidsonienne de l'interprétation : celui du locuteur et celui de l'interprète. La distinction entre ces deux points de vue est généralement à l'origine de la distinction que Davidson fait entre la connaissance par le locuteur d'une théorie-T pour son langage d'une part et à l'assignation charitable par l'interprète de la meilleure théorie-T qui donne un sens à ses jugements, effectuée de l'extérieur et sur la base des données empiriques. Mais Dummett nie que Davidson puisse faire cette distinction : il est obligé de confondre le point de vue de l'interprète et celui du locuteur, s'il veut expliquer comment l'assignation la plus charitable coïncide avec ce que *comprend* le locuteur. Or selon Dummett les deux composantes de la compréhension ne peuvent pas être simultanément attribuées au locuteur. Supposons en effet que l'interprète assigne par exemple à un nom « N » employé par le locuteur un référent sous une assignation charitable. Mais si le second composant de la compréhension (la connaissance du fait qu'une théorie-T permet de dériver des phrases-T vraies contenant « N ») fait partie de la compréhen-

1. Dummett, 1975 : 133.

sion qu'a le locuteur de « N », alors celui-ci ne peut pas donner une particularisation supplémentaire de la référence de « N », et ne peut donc considérer que ce qu'il sait fait partie de la signification de ce terme. En d'autres termes, le holisme l'empêche de connaître le sens des termes de *son* propre langage[1].

d / Dummett ajoute à ces critiques deux corollaires. Le holisme de Davidson, selon Dummett, n'est pas seulement constitutif ; c'est aussi un holisme épistémologique ou « inextricable » : il ne fait pas de distinction entre les phrases théoriques et les autres, et est donc incapable de distinguer un désaccord quant aux *significations* d'un désaccord quant à la vérité, et ainsi donner une place à la notion d'erreur[2]. Selon une théorie holiste, il devrait être toujours possible, quand on a des raisons de penser que deux individus ont des croyances divergentes, de penser que leurs divergences ne viennent pas de leurs croyances, mais de la signification de leurs termes. Selon Dummett, l'usage généralisé du principe de charité et l'incapacité de la théorie vériconditionnelle de distinguer le fait de tenir une phrase pour vraie du fait d'avoir des justifications de sa vérité doivent nécessairement conduire à cette conséquence. Enfin, un dernier corollaire d'une théorie modeste est qu'elle est incapable de rendre compte de l'apprentissage du langage. Si comprendre une phrase c'est comprendre le réseau entier d'un langage, on ne voit pas comment un langage pourrait jamais être appris. Le fait que Davidson insiste sur le caractère fini d'une théorie-T et sur sa récursivité comme condition d'apprentissage est insuffisant : car si on ne peut saisir le contenu d'un axiome individuel qu'en saisissant la structure totale de la théorie déductive dont il fait partie, le caractère récursif et moléculaire de la théorie ne suffit pas pour rendre l'apprentissage progressif ; au contraire la compréhension doit être d'emblée totale. Le molécularisme de principe des théories-T est donc miné par les contraintes holistiques de l'interprétation. Seul un molécularisme substantiel, admettant l'idée d'un apprentissage progressif des termes d'un langage et l'assignation isolée de significations à ses expressions, peut satisfaire à la condition d'apprentissage[3].

1. Dummett, 1987 : 244-245.
2. Dummett, 1975 : 117-120.
3. Dummett, 1975 : 137 ; 1978 : 302-303.

Il est frappant de constater, comme je l'ai déjà signalé, que l'ensemble de cette argumentation contre Davidson n'est pas dirigée contre son réalisme, mais contre son holisme. Mais il n'est pas difficile de voir en quoi le premier doit découler du second. Une théorie comme celle de Davidson est, selon Dummett, forcée de représenter la connaissance des conditions de vérité comme une connaissance de conditions transcendantes par rapport à leur vérification directe, puisque l'établissement de ces conditions doit s'effectuer sur la majorité des assignations charitables de signification. On pourrait certes parler néanmoins de vérification indirecte ou holistique. Mais celle-ci ne peut pas satisfaire aux conditions (A)-(M) énoncées ci-dessus. Un locuteur individuel, selon l'image holiste, n'a qu'une connaissance très indirecte, médiate, et jamais manifestable des sens de ses expressions : il doit sans cesse déférer cette connaissance à une assignation lointaine, qu'il suppose garantie par la charité, mais qu'il n'a jamais de raison de supposer telle, des conditions de vérité à l'ensemble du langage. *A fortiori* il devient impossible de rattacher cette connaissance à des capacités pratiques de recognition des expressions ni à des conditions d'assertion des phrases. C'est pourquoi Dummett finit par assimiler holisme et réalisme.

4.6. De l'importance d'être modeste

Les critiques de Dummett me paraissent contenir certaines des objections les plus sérieuses que l'on puisse adresser à Davidson. Mais plusieurs reposent sur des pétitions de principe, et d'autres sur des erreurs sur la position réelle de Davidson. Je voudrais ici faire provisoirement le point, en reprenant les critiques a /-d / de la section précédente.

a / Une théorie de la signification de type davidsonien présuppose-t-elle la compréhension du langage et est-elle en ce sens circulaire ou triviale ? Cela dépend dans une large mesure de ce qu'il faut entendre par « connaissance d'un langage », et de la relation qu'il y a entre celle-ci et les théories-T supposées l'éclairer. C'est un problème fondamental, sur lequel je reviendrai dans les chapitres suivants, mais dont on peut esquisser ici la formulation. A la différence de Dummett, Davidson se place

en général du point de vue de l'interprète, c'est-à-dire du point de vue des *attributions* de signification, et non pas du point de vue du locuteur. En ce sens une « théorie de la signification » est une théorie dont disposerait idéalement l'interprète. Tout le problème est de savoir si une telle théorie peut être attribuée à un locuteur réel et caractériser sa compétence sémantique effective. Dummett a raison de voir là une ambiguïté. Car d'un côté Davidson semble dire qu'une TS caractérise, de manière idéale, ce en quoi *doit* consister la connaissance d'un locuteur réel, sans pour autant soutenir que ce dernier connaît *en fait* quelque chose comme une théorie-T pour son langage. Mais d'un autre côté, Davidson considère aussi qu'il est légitime d'attribuer au locuteur lui-même une connaissance qui a au moins la *forme* d'une théorie-T pour son langage. Comme on l'a vu, pour Davidson, comprendre c'est interpréter, c'est-à-dire posséder, d'une manière ou d'une autre, une certaine sorte de *théorie*. Le locuteur apprend et comprend son langage en devenant interprète de ce que disent les autres, et de ce qu'il dit lui-même. En ce sens, il doit y avoir un lien entre ce que connaît le locuteur et la manière dont un interprète caractérise, de l'extérieur, ses significations et ses croyances en projetant sur son discours la structure d'une théorie-T. Dummett a raison de voir là une difficulté. Car la conception davidsonienne est vouée à attribuer au locuteur une compétence sémantique qui sera avant tout un trait de la théorie de l'interprète, et non pas de ce que *comprend* le locuteur lui-même, y compris dans le cas où le locuteur est *son* propre interprète. En ce sens, la caractérisation de la compétence sémantique que fournit Davidson devra nécessairement être externe à la manière dont le locuteur lui-même comprend son langage, et en ce sens indirecte, alors que pour Dummett seule une connaissance directe par le locuteur de ses propres significations peut être source de compréhension. Il y aura donc une *asymétrie* entre la connaissance qu'il a, à la première personne, de ses significations, et la connaissance qu'un interprète (y compris, encore une fois, lui-même) lui attribue à la troisième personne. Dummett semble soutenir qu'en raison de cette asymétrie, il y aura toujours un décalage entre la manière dont une « théorie » caractérise la compétence et la compétence effective du locuteur. Ou cette asymétrie est, comme le suggère Dummett, illégitime parce qu'elle interdit de caractériser véritablement la compréhen-

sion du langage, ou elle est légitime, mais il faut expliquer en quoi. On notera pourtant que cette ligne d'argumentation risque de mettre en cause un trait de la compréhension du langage sur lequel Dummett insiste lui-même, à savoir son caractère public et manifeste. Comment peut-on soutenir que l'asymétrie en question menace la compréhension qu'a un locuteur de *son* langage, et insister aussi sur la publicité des significations et sur la nécessité du point de vue « externe » propre à une théorie substantielle ? La position de Dummett semble donc elle aussi soumise à des tensions.

Ce point, sur lequel nous aurons à revenir abondamment, me paraît indépendant de la cible explicite de l'objection a /, qui impute une circularité à la représentation davidsonienne de la compréhension du langage. Il est exact que la TS forgée par un interprète est fournie sous l'hypothèse que le locuteur connaît son propre langage. C'est par exemple une hypothèse que l'interprète ne ferait pas, ou plutôt ne ferait pas de la même manière s'il avait affaire à un enfant, ou à un locuteur parlant une langue étrangère à sa langue maternelle. Mais cela ne veut pas dire qu'une TS de type davidsonien soit un simple exercice de représentation d'une connaissance qui serait déjà déterminée par ailleurs. Ce qui induit Dummett à dire que l'application d'une TS de ce type suppose une compréhension préalable du métalangage de la théorie (ou du fait que l'on saurait déjà que les phrases de droite des biconditionnels (T) sont des traductions des phrases de gauche) est le fait que les phrases-T sont conformes au schéma tarskien d'équivalence, qui présuppose une explication préalable du concept de vérité. Dummett en conclut que puisque la vérité est présupposée, et que la signification est déterminée par les conditions de vérité, la signification est également présupposée. Mais il ne s'ensuit rien de tel. Si Davidson prend bien le concept de vérité comme primitif et déjà compris, il n'affirme pas l'identité de la signification et des conditions de vérité, et encore moins que la *signification* soit déjà comprise, puisque l'interprétation radicale doit la déterminer. Peut-on dire alors qu'une TS est aussi inopérante qu'un manuel de traduction ? L'objection de Dummett repose sur une confusion qui mérite d'être rappelée (§ 1.3). Un manuel de traduction est inadéquat parce qu'il résulte de phrases de la forme « 'S' signifie (dans L) ce que 'S'' signifie (dans L') » où « S » et « S' » sont

mentionnées. Mais une TS homophonique conforme au schéma (T) ne mentionne pas la phrase du métalangage ; elle l'*utilise*. C'est pourquoi, comme on l'a vu, les phrases-T ne sont pas triviales. Il est vrai que les phrases utilisées appartiennent au métalangage de l'*interprète*, qui les assigne au locuteur, ce qui suppose bien que l'interprète comprenne ce métalangage, pour pouvoir ensuite, par approximations, déterminer s'il partage celui-ci avec le locuteur. Mais cela ne présuppose en rien que le locuteur partage ce métalangage[1]. D'une manière générale, cette confusion provient de la confusion systématique que fait Dummett entre le schéma tarskien (T), la thèse d'équivalence, et la théorie de la vérité-redondance. Comme Dummett le voit bien, la théorie de la vérité-redondance n'est applicable qu'à des énoncés déjà individualisés du point de vue de leur contenu ou de leur signification. Mais il n'est pas évident que la *phrase* « La neige est blanche » ait le même contenu que le *jugement* que *la neige est blanche*, bien que « La neige est blanche » soit vrai si et seulement si la neige est blanche[2]. Cette dernière équivalence est celle qui fonde la conception décitationnelle de la vérité, et les phrases-T d'une théorie homophonique. Or Davidson n'accepte pas la première, qui repose sur le concept de traduction, et utilise les secondes pour donner une assise empirique à une théorie-T. Dummett fait comme si la thèse d'équivalence, sous sa forme tarskienne, présupposait, comme la théorie de la vérité-redondance, non seulement une explication du concept de vérité, mais aussi les contenus des jugements ou leurs significations.

Enfin, est-il légitime d'assimiler une théorie modeste à la conception psychologiste du langage comme code des pensées ? Dans la mesure où Dummett tire cette conséquence de sa critique précédente, et que cette dernière est illégitime, la conséquence l'est aussi. Mais elle est également surprenante, puisque l'interprétation radicale selon Davidson ne présuppose pas une individuation des contenus de croyances antérieurement à une individuation des significations[3]. Il reste évidemment possible à Dummett de voir une forme de « psychologisme » dans l'idée que les

1. Evans & Mc Dowell, 1976 : x ; Mc Dowell, 1987, reprend précisément cette objection contre Dummett.

2. Baldwin, 1991 : 22.

3. C'est un point que je n'avais pas vu dans ma réponse à Dummett 1986 (Engel 1986).

significations et les croyances sont interdépendantes. Mais s'il s'agit d'une forme de psychologisme, il faut encore montrer en quoi, et en quoi il est illégitime, ce que Dummett ne fait pas.

b / Une théorie « modeste » est-elle holiste ? Ce holisme est-il constitutif ou radical ? Et interdit-il toute théorie de la signification digne de ce nom ? C'est évidemment un point fondamental, sur lequel les critiques de Dummett rejoignent d'autres critiques de Davidson[1]. J'y reviens plus loin (§ 6.5). Mais on doit déjà faire les remarques suivantes. Dummett note bien, comme on l'a vu, que la conception davidsonienne n'est que partiellement holiste. Ce qu'il appelle la première composante d'une théorie de la signification, une théorie-T, est bien molécullariste, puisqu'une théorie récursive détermine la structure compositionnelle des phrases sur la base de leurs éléments. En particulier une théorie-T est moléculaire dans son traitement des constantes logiques. Comme le dit Tennant, qui est un partisan de l'antiréalisme, « le davidsonien est un molécullariste en ce qui concerne le noyau d'une théorie de la signification quand il s'agit des clauses récursives de la théorie constituant le noyau d'une théorie de la signification »[2]. Dummett est prêt à accepter ce point, mais il soutient que dans l'application de la structure moléculaire à l'interprétation, le bénéfice de ce molécularisme est perdu, parce que c'est la structure d'ensemble du langage qui doit déterminer les assignations de signification. En d'autres termes, les TS davidsoniennes ne sont pas molécullaristes au sens plus fort requis par Dummett : le sens des phrases doit être déterminé par des éléments simples dont la signification doit être indépendante d'éléments plus complexes et d'autres phrases plus complexes. Cela revient, comme on l'a vu, à exiger une forme de conservativité des assignations de signification. La portée de cette critique dépend alors de la plausibilité de l'adoption d'un réquisit aussi fort pour une théorie de la signification, d'autant plus qu'il doit avoir pour conséquence l'adoption d'une conception intuitionniste de l'inférence logique.

c / Dummett ne voit rien à objecter non plus au holisme tant que celui-ci demeure méthodologique. Mais il conteste qu'il puisse rester *seulement* méthodologique, c'est-à-dire ne toucher que la façon dont les TS sont confirmées et pas la nature intrinsèque des significations. La première question est de savoir si ce holisme est plus que méthodologique. S'il l'est, une seconde question est de savoir s'il a bien les conséquences dirimantes que Dummett en tire. Et s'il a ces conséquences, une troisième question est de savoir si cette doctrine est aussi inacceptable que Dummett le soutient. On peut se demander si cette dernière objection ne présuppose pas ce qui est en question, c'est-à-dire qu'une théorie de la signification doit être substantielle et moléculaire. Je la laisserai à nouveau de côté pour le moment, pour répondre aux deux premières.

1. En particulier Fodor et Le Pore, 1992.
2. Tennant, 1987, chapitre 8 : 67 ; 1987a. Cf. Engel, 1989, chapitre XII.

Tennant soutient qu'il faut séparer les deux composantes d'une théorie davidsonienne de la signification — la théorie-T moléculaire et la procédure d'interprétation holistique — et que le holisme de la seconde est purement méthodologique. Dummett conteste cela, et soutient au contraire que le holisme est « interne » à la théorie, ce qui rend les deux composantes inconciliables[1]. Cette question est évidemment essentielle pour qui souhaite, comme Tennant, concilier une conception davidsonienne de l'interprétation avec des réquisits antiréalistes (§ 5.3 *infra*). J'ai plus haut (§ 1.2 ; § 2.3.1) décrit le holisme de Davidson comme reposant sur deux thèses fondamentales : 1 / le fait qu'il n'est pas possible de déterminer les significations des phrases d'un locuteur indépendamment de ses croyances et de ses actions et *vice versa*, et 2 / le fait qu'il n'est pas possible de déterminer un contenu de croyance ou de signification individuel indépendamment des contenus d'*autres* croyances ou phrases[2]. Je m'accorde donc avec McDowell :

> Tel que je le comprends, le holisme que Davidson accepte, c'est la thèse selon laquelle les attributions de contenus à des phrases dans le langage d'une communauté, leurs actes linguistiques, et leurs états psychologiques, sont systématiquement enchevêtrés [*interlocking*] de telle manière qu'... il n'y ait pas moyen d'expliquer le contenu en général, ou l'attribution d'un contenu spécifique à un terme particulier « de l'extérieur » du système entier. Cela implique clairement

1. Tennant, 1987 : 67 ; 1987a ; Dummett, 1987.
2. Le lecteur attentif aura remarqué que cette formulation diffère cependant de celle donnée au § 1.2, où il est dit que ce sont *toutes* les autres significations (et croyances) qui doivent déterminer une signification individuelle. Ce point est crucial. J'y reviens.

une répudiation de toute aspiration à être autre que modeste dans une théorie de la signification (McDowell, 1987).

Rien de ceci ne prouve que le holisme de Davidson soit plus que méthodologique, et que cela fasse partie de la *nature* des significations qu'elles soient « enchevêtrées » avec des croyances et des actions, et avec d'autres significations, croyances et actions. Du fait que les attributions de contenus soient « enchevêtrées » aux actes de langage des membres d'une communauté et à leurs attitudes, il ne s'ensuit pas que l'on ne puisse pas distinguer le point de vue de l'interprète de celui du locuteur. Le fait que, comme le dit Dummett, les données chargées de tester la théorie de l'interprétation soient « internes » à cette théorie n'implique pas que les assignations maximales charitables que l'interprète fait aux phrases d'un locuteur doivent faire partie de la compétence sémantique de ce locuteur, pas plus que, comme on l'a vu, le fait que l'interprète caractérise cette compétence sous la forme d'une certaine théorie n'implique que le locuteur comprenne son langage sous cette forme. Davidson doit certes exiger qu'il y ait un certain isomorphisme entre la théorie de l'interprète et la compétence de l'interprété, mais il n'a pas besoin d'exiger l'*identité* des deux. Dummett trouve le holisme de Davidson intenable parce que, selon ce modèle, un locuteur devrait posséder la totalité, ou une vaste majorité, des assignations charitables pour déterminer les conditions de vérité et de référence des expressions de son langage. Si c'était vrai, ce modèle manquerait en effet de toute plausibilité. Mais Dummett confond deux choses : les principes normatifs qui gouvernent la procédure d'interprétation, et l'application de cette procédure. Que la charité soit une norme de l'interprétation signifie que c'est une idéalisation dont l'interprète se sert pour déterminer les contenus ; cela ne signifie pas que l'on doit attribuer à un locuteur une compétence plus grande que celle qu'il possède en fait. C'est à nouveau (cf. § 3.5) prendre le principe de charité pour un principe descriptif, et confondre l'explication *à l'intérieur* de la théorie et explication *de* la théorie. A cela Dummett pourrait répondre que cela n'invalide pas son diagnostic selon lequel le locuteur a seulement une connaissance *suffisante* de la signification des termes de son langage, selon le modèle wittgensteinien pour le nom « Moïse ». Que l'on se place

du point de vue du locuteur ou de celui de l'interprète, rien ne garantit qu'ils parviennent aux mêmes assignations de vérité. Implicitement, cette critique revient à rejeter la thèse de l'indétermination de la traduction et de l'inscrutabilité de la référence, que Dummett attaque ailleurs (1978) directement. Et même si Davidson admet, comme on l'a vu, ces thèses, il soutient que les conséquences de l'indétermination de l'interprétation sont loin d'être aussi dramatiques. Il y a, sans aucun doute, un lien étroit entre le holisme et les thèses d'indétermination. Mais il reste à montrer en quoi ces principes détruisent toute théorie de la signification systématique.

d / Les mêmes remarques s'appliquent aux objections portant sur l'impossibilité pour une théorie holiste de distinguer des désaccords factuels de désaccords substantiels et de donner une conception plausible de l'apprentissage du langage. On dit souvent que c'est une des conséquences du rejet de la distinction analytique/synthétique que les questions portant sur la signification deviennent indistinguables de celles qui portent sur les faits. Mais à nouveau, une chose est l'affirmation de principe de l'« enchevêtrement » des croyances et des significations, et une autre la pratique de l'interprétation. Nous pourrions reprendre ici l'exemple de Mrs. Malaprop (§ 3.4). Mais on peut en prendre un autre (Davidson, 1974 : 196, *286*). Quelqu'un dit : « Regardez ce joli yawl ! » A-t-il confondu un ketch avec un yawl, ou bien devons-nous réinterpréter ses mots ? La charité nous enjoint de choisir la seconde solution. Mais cela n'implique en rien qu'un interprète conservera cette solution, ni qu'il ne pourra discerner une erreur factuelle. Il serait tout aussi absurde de soutenir la proposition inverse selon laquelle on devrait conserver systématiquement le sens des mots dans ce genre de cas et attribuer des erreurs factuelles. Davidson fait appel précisément au principe de charité pour expliquer comment l'erreur est possible (§ 2.3). Dummett semble objecter qu'être *trop* charitable ferait perdre son sens à la notion d'erreur. Mais il ne s'agit pas d'être trop ou pas assez charitable : la charité nous est imposée, ensuite l'interprète « triangule » entre ses jugements de vérité, ceux de l'interprété, et leur interaction avec le monde extérieur. Dummett a raison en ceci que l'on est en droit de se demander quel sens la position de Davidson peut donner

à la notion d'objectivité. Mais rien n'indique *a priori* que ce soit impossible[1]. Enfin, l'objection selon laquelle le holisme interdit l'apprentissage du langage est la plus faible. On ne voit pas pourquoi le holisme reviendrait à la thèse absurde selon laquelle celui qui apprend (au sens non innéiste du terme) un langage devrait apprendre le langage tout entier. Même si le holisme était la thèse « du réseau » que Dummett stigmatise selon laquelle le sens d'*une* expression repose sur le sens de *toutes* les autres, on ne voit pas pourquoi il serait impossible d'apprendre, fragment après fragment, un savoir dont la nature est systématique et globale. Le holisme ne propose aucune hypothèse empirique sur l'apprentissage. Davidson inclut bien la condition de finitude (§ 1.2) pour expliquer comment l'apprentissage est possible, mais il n'entend pas en décrire le processus. On peut à nouveau retourner les tables : le molécularisme d'une théorie substantielle n'entraîne-t-il pas une conception très restrictive de l'apprentissage, et n'est-ce pas parce que Dummett exige de toute théorie de la signification qu'elle réponde à ces critères qu'il dresse un constat d'impossibilité pour toute théorie non conforme ?

Ce qu'il faut donc retenir des critiques de Dummett me paraît se ramener à trois points principaux : la question de la nature de la compréhension d'un langage, celle de la nature du holisme, et celle de la nature du réalisme. Dans chacun des cas, nous pouvons tenir les doutes dummettiens comme légitimes. Mais il y a loin de là à l'acceptation du programme antiréaliste. Je reprendrai ces questions dans les chapitres qui suivent dans l'ordre inverse de celui où je viens de les énumérer, en commençant par décrire une position de type réaliste qui me paraît permettre de répondre aux doutes antiréalistes.

1. Cf. également sur ces points, les remarques de Laurier, 1991 : 125-127.

5

Le réalisme minimal

Tout est mini dans notre vie.
Jacques Dutronc.

5.1. Objectivité de la signification et objectivité de la vérité

L'argument antiréaliste standard établit une relation étroite entre la thèse selon laquelle comprendre la signification d'une phrase c'est saisir ses conditions d'assertion et la thèse selon laquelle la vérité doit elle-même être définie en termes d'assertabilité. Appelons, à la suite de John Skorupski (1988, 1992), la première thèse « conception épistémique » de la signification, et la seconde « conception épistémique de la vérité ». Les deux thèses en effet reviennent à imposer une condition épistémique sur les notions de signification et de vérité : l'antiréaliste soutient que la signification n'est pas une donnée purement objective qui existerait indépendamment de la *connaissance* qu'en ont les locuteurs, et il soutient que la vérité n'existe pas indépendamment de la *connaissance* que nous en avons, ou des raisons ou justifications que nous avons pour l'affirmer. La seconde thèse est évidemment ce qui rattache l'antiréalisme dummettien à l'idéalisme ou au vérificationnisme traditionnels. La première est ce qui fait l'originalité de cet antiréalisme, dans la mesure où il prend explicitement la forme d'une conception de la signification. S'il est vrai que l'idéalisme traditionnel — berkeleyien par exemple — s'appuie en grande partie sur des considérations sur le sens des mots pour « rejeter » la « réalité » du monde extérieur, il reste néanmoins un réalisme

sémantique, au sens où les énoncés portant sur des objets matériels sont supposés pouvoir se réduire à des énoncés portant sur des données sensorielles, qui ont des conditions de vérité (mentalistes)[1]. Au contraire l'antiréalisme dummettien s'en prend directement à cette association entre signification et conditions de vérité, et sa critique de cette association n'implique en rien un idéalisme au sens traditionnel : on peut être antiréaliste en sémantique sans rejeter la « réalité du monde extérieur » et sans souscrire à une forme d'instrumentalisme en philosophie des sciences. L'antiréaliste dummettien n'a pas à nier que nous fassions partie d'un monde objectif et public, que nous n'avons pas inventé[2]. Pour reprendre une distinction voisine de la précédente, proposée par Wright (1987 : 5-13), on ne doit pas confondre *l'objectivité de la signification*, d'après laquelle la signification est une condition réelle, indépendante de nous, *l'objectivité de la vérité*, selon laquelle la vérité est indépendante de notre connaissance, et *l'objectivité du jugement*, selon laquelle nos énoncés sont capables d'enregister des traits du monde réel, indépendamment de nos capacités épistémiques. Les débats antiréalistes portent sur la relation entre les deux premiers types d'objectivité, mais n'impliquent rien quant au troisième[3]. Si nous faisons ces distinctions, l'argument antiréaliste standard apparaît à la fois plus fort, puisqu'il soutient que le rejet de l'objectivité de la signification et la conception épistémique de la signification impliquent le rejet de l'objectivité de la vérité et l'adoption de la conception épistémique de la vérité, et plus faible, parce que cette implication n'a rien d'évident. Il *semble* y avoir un lien naturel entre l'idée que la signification n'est pas constituée par des conditions de vérité « transcendantes » ou « indétectables » mais par des conditions d'assertion et l'idée que la vérité elle-même se définisse comme assertabilité. De fait la seconde thèse implique la première : si la vérité est l'assertabilité, alors les conditions de vérité doivent être des conditions d'assertion. De même l'objectivité de la vérité implique celle de la signification : si tous nos énoncés sont vrais « indétectablement », alors leur sens est lui-même objectif. Mais les implications converses ne valent pas. On peut nier l'objectivité de la signification, comme Quine par exemple, sans rejeter l'objectivité de la vérité[1]. Et surtout on peut admettre la conception épistémique de la signification sans admettre la conception épistémique de la vérité. Du fait que la signification soit « épistémiquement contrainte », et puisse s'analyser en termes de conditions d'assertion, il ne s'ensuit pas que la notion de vérité ne soit pas objective, et que la vérité doive se définir comme l'assertabilité ou la justification garantie, au sens vérificationniste. Ou du moins c'est ce que je voudrais essayer de montrer dans ce chapitre. S'il est possible de soutenir que l'on peut admettre toutes, ou certaines, des conditions que Dummett exige d'une théorie de la signification *sans* admettre les conséquences qu'il en tire pour la vérité, la révision de la logique classique au profit d'une logique intuitionniste, et si l'on peut soutenir aussi que certaines conditions intuitionnistes sont compatibles avec une théorie de la signification davidsonienne, la quasi-équivalence présupposée par l'antiréaliste entre l'objectivité de la signification et l'objectivité de la vérité apparaîtra douteuse. Non seulement l'argument antiréaliste standard s'en trouvera profondément affaibli, mais aussi la différence entre la position qualifiée de « réaliste » et la position « antiréaliste » en sera diminuée d'autant. Elle devrait, si ceci est correct, se ramener à la distinction entre une conception épistémique et une conception non épistémique *de la signification*. Mais s'il est vrai également qu'une conception davidsonienne impose bien certaines contraintes épistémiques sur la signification, alors le débat entre le réalisme et l'antiréalisme, tel qu'il a été présenté jusqu'ici, perdra en partie sa substance. Mon but n'est pas de défendre ces points à des fins dialectiques ou de réfutation seulement, pour dire que la question du réalisme n'a en fait rien à voir avec une réflexion sur la nature de la signification, ou pour soutenir à l'inverse qu'une réflexion sur la signification ne nous donne aucun accès à des propositions sur le réalisme. Je voudrais plutôt soutenir qu'il existe une sorte de région intermédiaire, où l'on peut nouer les liens entre les deux questions, et défendre ce que

1. C'est, comme on l'a vu, ce que Dummett appelle un réalisme « sophistiqué » cf. les références de la note 19, chap. 4

2. Cf. Tennant 1987, chap. 2, qui critique sur ce point les analyses de Devitt, 1984, qui tend à confondre systématiquement ces thèses.

3. Cf. Engel, 1993.

1. La thèse de l'indétermination n'est pas incompatible avec le physicalisme. Ou du moins c'est ce que soutient Quine.

l'on peut appeler une forme de réalisme *minimal* de la vérité et de la signification, qui rejette les implications les plus tranchées des thèses « réalistes » et « antiréalistes » tout en conservant ce qu'elles ont de correct[1]. L'idée, pour la présenter brièvement, est que notre compréhension intuitive du concept de vérité repose sur des platitudes, telles que « asserter que *p*, c'est dire que *p* est vrai », ou « une phrase est vraie en vertu des faits ». La position courante des philosophes qui discutent de l'analyse de la notion de vérité[2] est que ces platitudes ne sont pas innocentes ou anodines, et qu'elles enveloppent des présuppositions métaphysiques profondes, qu'il convient de défendre ou de rejeter. Mais il existe une autre position, selon laquelle ces platitudes ne sont *que* des platitudes, qui expriment l'essentiel de ce que l'on doit entendre par « vérité », qui n'impliquent ni n'endossent aucune thèse métaphysique sujette à controverse. C'est une position que l'on a pu appeler diversement *neutraliste, déflationniste, quiétiste,* ou *minimaliste*[3]. Nous allons voir qu'il y a diverses versions de cette thèse. En particulier on doit distinguer la version selon laquelle *il n'y a rien d'autre à dire sur la vérité, métaphysiquement parlant, que ce que contiennent ces platitudes*, parce que la vérité n'est pas une propriété « substantielle » (déflationnisme et quiétisme), de la version selon laquelle *bien que les platitudes représentent le sens du prédicat « est vrai », elles ont un certain import métaphysique*, parce qu'elles disent quelque chose de substantiel sur la vérité (minimalisme). Ce quelque chose de substantiel exprime une certaine forme de réalisme minimal, qui est une thèse métaphysique significative. Si on devait s'en tenir aux platitudes, le débat dummettien entre le réalisme et l'antiréalisme n'aurait pas de sens, alors que pour

1. J'ai déjà défendu ce « réalisme minimal » quant aux vérités logiques dans Engel, 1989 (l'expression n'apparaît que dans la version de 1991, celle de 1989 parlant de « conventionnalisme minimal »). Mais il y a, comme on va le voir, plusieurs versions du minimalisme quant à la vérité, quant à la signification. Les versions les plus proches de celle défendue ici sont celles celle de Peacocke (1986) et de Wright (1987, 1993).

2. Je veux dire des philosophes qui se posent la question de savoir ce que signifie le mot « vrai », pas celle de ceux qui se contentent de présupposer que la vérité s'identifie à telle ou telle notion, comme la correspondance. Cf., outre les références de la note suivante, Ramsey, 1926 ; Wittgenstein ; Dummett 1959, 1978 ; Davidson, 1969.

3. Cf. Blackburn, 1984 ; Field, 1986, Baldwin, 1987, Mc Dowell, 1987 ; Horwich, 1990 ; Wright, 1993, et Engel, 1989, chap. 5. Certains auteurs « néo-pragmatistes », ou se disant tels, comme Rorty, défendent aussi une position de ce genre. cf. *infra* 6.4.

le minimaliste, il a un sens, bien que s'ils s'accordent sur la signification minimale de la notion de vérité, le réaliste et l'antiréaliste sont voués à s'accorder sur un certain nombre de points cruciaux. Ces nuances peuvent, de prime abord, paraître assez vaines, mais si elles ont un sens, le débat engagé par Dummett prendra une autre tournure.

5.2. Les nombreuses facettes du réalisme

Revenons à la caractérisation du réalisme de Dummett. Il consiste en :

(I) la conception véri conditionnelle de la signification :
  (i) le concept central d'une théorie de la signification est le concept de vérité
  (ii) ce concept est expliqué en termes du schéma véri conditionnel (T) « S est vrai ssi *p* »
  (iii) notre connaissance de la signification d'une phrase S consiste dans la connaissance de ses conditions de vérité sous la forme du schéma (T)
(II) la transcendance de la vérité par rapport à la vérification
(III) l'admission du principe de Bivalence pour les énoncés non décidables
(IV) pour toute phrase, il doit y avoir quelque chose en vertu de quoi elle est vraie (vérité correspondance)

Dummett, comme on l'a vu, présente souvent ces thèses comme si elles s'impliquaient mutuellement, ou étaient équivalentes. Il les présente aussi parfois (dans ses premiers écrits) comme si elles découlaient toutes de (III), le principe de Bivalence. Il est plus raisonnable de penser que le réalisme qu'attaque Dummett est la *conjonction* de ces thèses, et que l'abandon de l'une ou de plusieurs d'entre elles relâche l'engagement réaliste. Nous allons voir que ces thèses ne sont pas équivalentes. Notre tâche va consister à désintriquer les thèses les unes des autres[1]. Admettons cependant, pour le moment, qu'elles soient plus ou moins équivalentes, et laissons de côté la question de la Bivalence (III). Il existe une certaine sorte de théorie — ou d'image — réaliste qui semble bien correspondre à celle que Dummett attaque. C'est la conception que Putnam (1978, 1981) appelle *réalisme externe* ou *métaphysique* :

1. Dans tout ce paragraphe, je suis très redevable à Wright, 1987, en particulier chap. 10, à Wright 1993, ainsi qu'à Skorupski, 1988 et 1992.

> Le monde est constitué d'un ensemble fixe d'objets indépendamment de l'esprit. Il n'existe qu'une seule description de « comment sont les choses » (*the way the world is*). La vérité est une sorte de correspondance entre les mots ou des symboles de la pensée et des « choses » ou des ensembles de choses extérieures (Putnam, 1978 : 123).

Selon ce réalisme métaphysique, la vérité est radicalement indépendante de notre connaissance — tous nos énoncés pourraient être faux sans que nous le sachions — (thèse II) ; la réalité est elle-même composée de choses ou d'ensembles de choses (faits) indépendants de notre esprit ; nos énoncés sont vrais ou faux en vertu de leur correspondance avec ces faits (thèse IV). On peut se demander si tout réalisme externe doit reposer sur une conception correspondantiste de la vérité, c'est-à-dire si (II) implique (IV). Mais on supposera pour l'instant que c'est le cas. Nous dirons, dans le vocabulaire de la section précédente, que ce réalisme externe (RE) implique une conception totalement *non épistémique* de la vérité (NEV). Il y a bien des manières de critiquer RE. On peut, comme Goodman (1978), soutenir l'idée d'une pluralité possible de descriptions du monde et de « mondes », ou comme Putnam (1978) invoquer des arguments proches de celui de la relativité de l'ontologie de Quine (1969) pour montrer qu'il ne peut pas y avoir « une seule histoire vraie » sur *the way the world is*. On peut aussi critiquer la notion même de « faits » auxquels seraient supposés correspondre nos énoncés, comme Davidson (1969)[1]. Ce n'est pas, comme on l'a vu, la stratégie de Dummett, même si l'on peut supposer qu'il a peu de sympathie pour le réalisme externe. Ce qui intéresse Dummett est la relation entre NEV et la conception vériconditionnelle de la signification (I), selon laquelle connaître la signification d'une phrase, c'est connaître ses conditions de vérité. Il y a un lien d'implication évident entre NEV et (I) : si la vérité est radicalement non épistémique, et si connaître la signification d'une phrase c'est connaître ses conditions de vérité, alors connaître les conditions de vérité d'une phrase c'est connaître des conditions de vérité qui sont en principe indépendantes de notre

1. Cf. Engel, 1989, chap. V. Les critiques de l'atomisme logique de Russell et de Wittgenstein sont évidemment des critiques de ce type. Ce que Rorty (1979) appelle la conception représentationnelle de la réalité ou de nos énoncés comme « miroir » de la nature est également une image de ce type.

connaissance ou de notre vérification, « transcendantes » ou « indétectables ». Mais cette inférence *présuppose* NEV, et par conséquent n'intéresse pas Dummett s'il ne veut pas faire de pétition de principe contre le réalisme externe. Ce qui l'intéresse plutôt est le lien entre (I), la conception vériconditionnelle, et NEV. S'il peut montrer que (I) implique NEV, et que (I) est une conception indéfendable de la compréhension du sens, alors son argument aura réussi. Comme nous allons le voir, cette implication n'est pas du tout évidente, mais supposons provisoirement qu'elle le soit.

J'admettrai, comme Dummett, que le réalisme externe est faux (mais je ne donnerai pas d'arguments avant le chapitre suivant). Mais même si on acceptait cette thèse, cela ne changerait rien au problème. La question importante est de savoir si l'on peut critiquer (I) indépendamment d'une acceptation ou d'un rejet de NEV (entendu comme (II)-(IV)). Dummett soutient évidemment que oui, en vertu de l'argument antiréaliste standard : selon lui, (I) suppose au moins que le sens transcende l'usage, que la signification est une donnée totalement objective indépendante de la connaissance qu'en ont les locuteurs, et qu'elle consiste en des conditions de vérité indétectables. Cela revient à adopter une conception totalement objective, non épistémique, de la *signification* (et non pas, par conséquent, selon la ligne présente d'argumentation, de la *vérité*). Appelons-la (NES). Or la signification *est* sujette à des contraintes épistémiques sur sa compréhension : le sens doit être manifesté dans l'usage, acquis, et public (selon les principes (M) et (A) du 4.2). D'où la contraposition dummettienne. Mais si nous respectons les distinctions que nous venons de faire, tout ce que la contraposition en question nous permet d'affirmer n'est pas (comme (iii) au 4.1) que le *réalisme* en général (I)-(IV) est faux, mais que la conception non épistémique, vériconditionnelle, de la signification l'est. Il s'ensuivra qu'une conception épistémique de la signification, conçue en termes de conditions d'assertion, sera la conception correcte. Appelons-la (ES). C'est ce qui devrait suffire à l'argument dummettien. Le poids de cet argument devra donc reposer sur la critique de cette conception non épistémique.

A nouveau s'ouvrent deux voies d'argumentation qui peuvent permettre à Dummett de critiquer la conception vériconditionnelle (I) : l'une a / consiste à assumer la fausseté de NEV et la vérité d'une concep-

tion non épistémique de la vérité, pour en inférer ES, la conception épistémique de la signification ; l'autre b / consiste à inférer de ES une conception épistémique de la vérité (EV), selon laquelle la vérité doit être définie en termes d'assertabilité, et à soutenir que *puisque (I), la conception vériconditionnelle implique NEV, la négation de EV*, alors, par contraposition, (I) ou NES doit être fausse. Mais on peut voir immédiatement que ces deux voies ne permettent pas à Dummett de critiquer *directement* NES sans présupposer ce qui est en question, à savoir respectivement (EV) et (ES). On a déjà remarqué que, pour que l'argument antiréaliste standard soit opérant, Dummett doit accorder à son adversaire sa prémisse (I) ou NES, en montrant qu'elle a des conséquences inacceptables. Mais on a vu aussi qu'il était très difficile de considérer avec lui ces conséquences comme inacceptables sans admettre aussi la conception épistémique du sens et de la vérité qu'il propose. Beaucoup de critiques ont accusé ici Dummett de produire un argument circulaire[1]. Il est cependant encore trop tôt pour donner un diagnostic. S'il est vrai que, comme le veut l'adage, le *modus ponens* d'un philosophe est le *modus tollens* d'un autre philosophe, nous avons intérêt à nous concentrer autant sur les prémisses que sur les arguments eux-mêmes. Envisageons donc les deux voies (a) et (b).

On doit faire d'abord une remarque — qui s'avérera cruciale par la suite. Il semble, de prime abord, y avoir un lien direct entre assertion et vérité, qui s'exprime dans l'équivalence qui fonde le schéma tarskien (T) : asserter que *p*, c'est asserter qu'il est vrai que *p*. En ce sens, la signification de *p* c'est la condition d'assertion de *p*, et savoir que *p* est vraie, c'est savoir que *p* peut être assertée. Mais on ne saurait conclure de cette banalité que l'on peut tenir, comme faisant partie du sens usuel de « vrai », que EV et ES sont équivalentes. Car si cette équivalence nous dit bien que le *sens* de « vrai » et celle de « *peut être asserté* » est le même, elle n'implique pas que leurs *extensions* soient identiques (« vrai » veut dire « peut être asserté », mais tout ce qui peut être asserté n'est pas vrai).

La voie a /, qui conduit de EV à ES, est aisée, et je l'ai déjà indiquée ci-dessus. Supposons que nous admettions, pour une raison indépendante de ES, que la vérité est l'assertabilité garantie (« garantie », car toute asser-

1. C'est en partie l'une des leçons de notre analyse du chap. 4.6 ; cf. aussi Bilgrami, 1986.

tion n'est pas garantie ou correcte, et selon cette conception seules celles qui le sont ont des chances d'être *vraies*, mais je ne répéterai pas ce qualificatif). Alors les conditions de vérité d'une phrase consistent en ses conditions d'assertion, et par conséquent si comprendre la signification d'une phrase c'est comprendre ses conditions de vérité, alors comprendre la signification d'une phrase (comprendre cette phrase) c'est comprendre ses conditions d'assertion. Nous dirons que le *contenu* d'une phrase S est le même que celui de *il est assertable que S*. Cette série d'équations repose sur l'équation initiale, que l'on peut admettre, en particulier au nom d'une conception vérificationniste ou d'une conception intuitionniste de la vérité. Mais Dummett est très soucieux de ne pas présupposer une telle conception pour défendre son antiréalisme, et il refuse que celui-ci soit purement et simplement assimilé à une variante du positivisme viennois ou d'intuitionnisme (brouwerien ou autre), même si *une forme* de vérificationnisme ou de théorie épistémique de la vérité a sa faveur[1]. Par conséquent la voie a / n'est pas la bonne.

Essayons donc la voie b /, plus prometteuse parce qu'elle infère EV de ES. Mais elle est loin d'être aussi aisée que la précédente. Supposons que EV soit vraie, indépendamment de raisons que nous pourrions avoir en faveur de ES. Nous pourrons dire que puisque la signification d'une phrase est déterminée par ses conditions d'assertion, le contenu d'une assertion d'une phrase sera qu'elle *a* telles conditions d'assertion. Mais puisque, en vertu de l'équivalence banale ci-dessus, asserter que *p* c'est asserter que *p* est vrai, ou asserter que ses conditions de vérité sont réalisées, celles-ci doivent s'identifier à ses conditions d'assertion. On infère donc ainsi EV de ES.

Mais que l'on suive la première ou la seconde voie, la thèse antiréaliste, selon laquelle le contenu de *p* est le même que celui de *il est assertable que p*, n'est pas acceptable, pour la raison déjà mentionnée ci-dessus. Considérons par exemple leurs négations. *Il n'est pas vrai que p* et *il n'est pas assertable que p* n'ont pas les mêmes conditions d'assertion. Si *il n'est pas vrai que Gaston Dominici soit coupable* équivalait à *il n'est pas assertable que Gaston Dominici soit coupable*, de nombreuses affaires criminelles seraient vite résolues. « 'p' est vrai » et « 'p' est assertable » ne sont pas équiva-

1. Cf. par exemple Dummett, 1976 : 117.

lents dans les cas où « *p* » a des conditions d'assertion non concluantes. Plus généralement cela montre que le schéma tarskien

« S » est T ssi *p*

ne semble pas pouvoir être interprété au sens où « est T » signifie « est assertable ». C'est l'une des raisons pour lesquelles Dummett (1959) rejetait toute analyse de la vérité et de la signification fondée de ce schéma. Cela ne signe pas l'arrêt de mort de la stratégie antiréaliste, mais cela montre que l'antiréaliste ne peut pas, contrairement à ce qu'il souhaiterait, accepter la prémisse « réaliste » selon laquelle la signification consiste dans les conditions de vérité. Même si « 'p' est vrai » et « 'p' est assertable » ont le même sens, ils n'ont pas la même extension. L'hypothèse de l'équivalence entre *asserter que p* et *il est vrai que p*, qui fonde la théorie de la vérité comme redondance et la théorie décitationnelle n'entraîne pas l'équivalence de la vérité et de l'assertabilité (si elle l'entraînait, « vrai » ne serait pas, comme on va le voir, un prédicat « neutre »). Cela explique les hésitations de la stratégie de Dummett, qui en 1959 admettait cette thèse mais rejetait une conception vériconditionnelle de la signification, et en 1978 (*Preface*) rejetait la théorie de la vérité-redondance et admettait qu'une théorie de la signification formulée en termes des conditions de vérité puisse être interprétée comme une théorie des conditions d'assertabilité. Tirons-en, provisoirement, la conclusion qu'il n'est pas facile pour l'antiréaliste de soutenir purement et simplement ES sur la base d'une assimilation des conditions de vérité à des conditions d'assertion. Il lui faut élaborer cette dernière notion.

Y a-t-il, pour l'antiréaliste dummettien, des moyens indépendants de EV et de ES, entendues au sens ci-dessus, pour défendre une conception épistémique de la signification ? Oui, il y a, comme on l'a vu, les thèses d'acquisition (A) et de manifestation (M), et les thèses (R) et (C) (§ 4.A) de la réduction de la compréhension des expressions à des capacités de reconnaissance décidables. Le problème est que que si (A) et (M) peuvent paraître en général tout à fait acceptables (y compris, comme on va le voir, pour un « réaliste »), (R) et (C) sont beaucoup plus problématiques. Plusieurs critiques de Dummett, et en particulier McDowell, y ont

vu une forme de vérificationnisme béhavioriste, selon laquelle il faudrait chercher, *dans le comportement* observable des locuteurs, des marques fiables de ces capacités pratiques de recognition[1]. D'autres ont suggéré qu'il pouvait aussi s'appuyer sur une forme de cartésianisme, selon lequel les contenus de signification « directement » vérifiables étaient accessibles au locuteur seul, à la première personne[2]. Dans ce cas, la critique du réalisme serait que des conditions de vérité transcendantes ne peuvent pas être manifestées dans des capacités ainsi manifestables dans le comportement ou saisissables à la première personne. Il ne me semble pas sérieux d'attribuer à Dummett ce genre de positions. Mais il nous doit cependant un exposé plus clair de ce qu'il entend par « vérification directe », « moyen effectif de recognition » ou « manifestation complète du sens dans l'usage ». Quoi qu'il en soit, si la thèse de Dummett est que la signification consiste dans des aptitudes ou des dispositions pratiques de recognition canoniques manifestables dans « l'usage » ou peut-être le comportement, il doit soutenir que ces aptitudes existent et sont détectables *isolément*, une à une, s'il veut satisfaire aux réquisits molécularistes et à sa conception de l'apprentissage. Mais cette hypothèse se heurte à une objection familière, formulée par Wright lui-même :

> Il n'y a pas de comportement, ou de syndrome de comportement, qui puisse être considéré comme distinct de la possession par un sujet d'une croyance particulière. Les croyances qui conduisent typiquement à certaines conduites peuvent aussi bien trouver leur expression dans d'autres conduites, tout à fait opposées si nous imaginons des changements appropriés dans d'autres croyances d'arrière-plan du sujet et dans son réseau de désirs... De la même manière, le comportement linguistique qui exprime la compréhension qu'a quelqu'un d'un énoncé particulier sera une fonction non pas seulement de cette compréhension, mais aussi de ses autres croyances au sujet du monde et au sujet de ses auditeurs, ainsi que des ses intentions et croyances d'arrière-plan. La signification ne peut donc être manifestée exhaustivement dans l'« usage » (Wright, 1987 : 22).

1. Cf. en particulier Mc Ginn, 1980, Mc Dowell, 1981, Schiffer, 1987 : 224-227, Bilgrami, 1986 : 114-116. Mc Ginn compare notamment la contraposition dummettienne (§ 4-4.1) à celle que fait selon lui Quine : tous deux admettent que si le réalisme est vrai, la signification transcende l'usage ; mais alors que Quine détache (la signification transcende l'usage), Dummett contrapose.

2. Bilgrami, *ibid.* : 109 J'ai moi-même noté la possibilité en question ci-dessus au § 4.6.

Davidson ne trouverait rien à y redire. Wright soutient néanmoins que ce genre d'objection holiste ne porte que sur la *manifestation* des aptitudes à la compréhension, et non pas sur leur *nature* : « C'est entre l'aptitude et sa manifestation comportementale que, si l'on peut dire, tombe l'ombre du holisme » (*ibid.* : 23). Mais même si nous admettons cette distinction entre la nature des aptitudes ou des états qui constituent la compréhension, et leur manifestation publique ou comportementale, il faut admettre que l'important, pour l'antiréaliste dummettien, n'est pas la manifestation de la compréhension, mais sa nature. Et si l'important est la nature de la compréhension, en quoi l'argument qui porte sur sa *manifestation* doit-il compter (Skorupski, 1988) ? Nous avons déjà vu, au § 4.6, que l'« ombre du holisme » planait sur l'argument de la manifestation, et c'est bien pourquoi Dummett critique le holisme. Le poids de l'argument doit donc porter sur (R). Mais pourquoi devrait-on admettre (R) ? Et surtout, si la compréhension consiste dans des aptitudes pratiques, pourquoi le réaliste devrait-il rejeter cette idée, ou pourquoi, s'il l'accepte devrait-il soutenir que ces aptitudes sont des aptitudes à reconnaître des conditions de vérité indétectables ? C'est évidemment cette thèse qui paraît la plus invraisemblable : l'idée que l'on serait, pour ainsi dire en possession d'une sorte de photographie dont le contenu ne nous est jamais révélé, que l'on aurait une compréhension d'états de choses par nature hors de notre portée, d'un Grand Dehors, une Vue de Nulle Part. Mais pourquoi le réaliste devrait-il croire cela ? Il peut croire cela s'il suppose, à l'instar du réalisme externe, qu'il y a *plus* dans les conditions de vérité d'une phrase que ses conditions d'assertion, quelque chose qui les transcende et qui pour ainsi dire les sous-tendrait. Cela paraît en effet mystérieux, l'expression d'une croyance en une sorte de *Ding an Sich*[1]. Mais une chose est de soutenir que la vérité transcende totalement la vérification, et autre chose de soutenir qu'il doit exister une *différence* entre vérité et assertabilité. Et une chose est de soutenir la conception vériconditionnelle de la signification, et autre chose de soutenir le réalisme externe.

1. Cf. Blackburn, 1987 : 221, Blackburn, 1989.

5.3. *Déflationnisme et ses platitudes*

Nous avons noté que le raisonnement ci-dessus conduisant de ES à EV reposait sur l'équivalence entre « *p* » et « il est vrai que *p* », mais que celle-ci ne suffisait pas par elle-même à justifier la transition. Reconnaître le lien intrinsèque — l'identité en extension — entre affirmer qu'il est vrai que *p* et asserter que *p* ne nous engage pas à assimiler vérité et assertabilité. C'est une interprétation supplémentaire du sens de « est vrai » que nous ajoutons à cette platitude. Pareillement, il semble que nous puissions accepter la platitude en question, et interpréter le sens de « est vrai » comme « est transcendant par rapport à la vérification » et « correspond aux faits ». Or l'équivalence semble être ce sur quoi se fonde le schéma tarskien (T). Il s'ensuit apparemment que l'on peut admettre le lien entre vérité et assertion et la conception vériconditionnelle de la signification (I) *sans que cela nous engage à adopter soit EV, soit NEV*. En d'autres termes (I) semble neutre par rapport à l'une ou l'autre des conceptions « métaphysiques » de la vérité que nous avons considérées. Il faut distinguer ceci de la thèse parallèle : (I) *ne nous engage pas à accepter soit ES soit NES*, et donc neutre par rapport à l'adoption d'une théorie épistémique ou non épistémique de la signification. Mais l'idée que l'équivalence soit « neutre » vis-à-vis de ces diverses thèses se distingue d'une idée plus radicale : qu'il n'y ait *rien d'autre à dire* quant à la vérité et la signification que ce que contient l'équivalence et la conception vériconditionnelle. C'est cette idée qui fonde ce que l'on appelle une conception « déflationniste » ou « minimaliste » de la vérité et de la signification[1]. Cette conception ne

1. Il y a un problème de terminologie. Bien que l'idée soit présente dès Ramsey (et peut-être Frege, les scholastiques quand ils parlent de « transcendantaux, etc.), et chez Tarski, Wittgenstein, Ayer, Quine, etc., le terme « déflationnisme » est employé, à ma connaissance, d'abord par Field, 1987. Il est repris par Horwich, 1990, mais ce dernier appelle aussi sa position « minimaliste ». Le problème est encore compliqué par le fait que Johnston, 1988, emploie aussi ce dernier terme pour une conception de la signification, et que Wright, 1993, appelle également « minimalisme » une conception de la vérité qui n'est pas, selon lui, « déflationniste ». J'ai réservé ici le terme « déflationnisme » pour des conceptions de la vérité et de la signification qui soutiennent qu'il n'y a rien d'autre dans ces notions que les platitudes, et j'ai conservé le terme de « minimalisme » pour une position comme celle de Wright et comme les autres que je rangerai sous la rubrique du « réalisme minimal ». Dans ce qui suit, je néglige la question historique de savoir si Tarski était un déflationniste en ce sens. Elle est discutée par Davidson, 1990.

s'accorde pas avec la position de Dummett, pour qui adopter (I) c'est déjà mettre le doigt dans l'engrenage réaliste.

Considérons d'abord la conception déflationniste de la vérité (DV). Il y a en fait une famille de conceptions de ce genre, dont la théorie de la vérité-redondance et la théorie décitationnelle font partie. Je tiendrai ici la conception déflationniste comme étant celle selon laquelle le sens du prédicat « vrai » est entièrement fixé par le schéma

(T) « *p* » est vrai ssi *p*

qui fait de ce prédicat un dispositif de « décitation » et « S », et un dispositif d'assertion de S. Nous avons déjà vu pourquoi DV suppose que l'on connaisse la signification de S[1]. Je laisserai aussi de côté le problème posé par des contextes comme « Tout ce que dit le pape est vrai ». La thèse déflationniste soutient qu'il n'y a rien d'autre à dire sur la vérité, et donc que le prédicat « vrai » n'exprime aucune propriété « substantielle », en particulier aucune propriété métaphysique ou profonde, telle que la correspondance ou l'assertabilité garantie. Une conception déflationniste doit certes rendre compte de notre intuition selon laquelle la vérité est associée à l'assertion et à la correspondance, mais sans que cela implique que la vérité *se définisse* par l'assertabilité ou la correspondance. Le trait décitationnel indique la relation entre la vérité d'une phrase et son assertion. Mais comme nous l'avons vu au sujet du raisonnement (b) ci-dessus, on peut admettre cette platitude sans admettre la réduction de la vérité à l'assertabilité garantie. Qu'en est-il de la platitude portant sur l'association de la vérité à la correspondance :

(i) « *p* » est vrai ssi « *p* » correspond aux faits

derrière laquelle le théoricien correspondantiste est prêt à voir une propriété profonde de la notion de vérité ? Le déflationniste peut répondre que cela ne dit rien de plus que

(ii) « *p* » est vrai ssi les choses sont comme « *p* » dit qu'elles sont

1. Puisque l'équivalence entre la vérité de « *p* » et l'assertion que *p* présuppose que l'on comprenne le sens de « *p* », ou que dans le schéma tarskien la phrase de gauche traduise celle de droite. Cf. § 3.6 et 4.4.

En effet, chaque fois que l'on affirme

« *p* » dit que *p*

il s'ensuit, par (ii), que

Les choses sont comme « *p* » dit qu'elles sont ssi *p*

et en appliquant le schéma décitationnel (T) on retrouve la platitude (ii)[1]. Ou encore, supposons que nous exprimions l'intuition de la correspondance ainsi :

(iii) « *p* » est vrai *parce que p*.

Horwich (1990 : 111) suggère que nous nous donnions les lois physiques de l'univers, dont nous déduisons *p*. Il ne nous reste plus qu'à appliquer le prédicat « est vrai » (par citation, opération inverse de celle de décitation), pour obtenir

« *p* » est vrai.[2]

En d'autres termes, comme le remarque Davidson (1969, 41), des prédicats tels que « correspond aux faits », « c'est un fait que », ou « les choses sont telles que » ne disent rien de plus que « il est vrai que » ou « ... et c'est vrai ». Nous n'*ajoutons* donc rien au prédicat « est vrai » en le caractérisant au moyen de la notion de correspondance. (Aristote était peut être le premier déflationniste quand il disait que « dire ce qui est qu'il est, ou de ce qui n'est pas que ce n'est pas est vrai ».) Le déflationniste soutient que hormis la propriété formelle de décitation, « il n'y a rien de plus » dans la vérité qu'une série de platitudes :

Asserter que *p*, c'est dire que *p* est vrai
Nos énoncés sont vrais *en vertu* de quelque chose (les faits)

1. Wright, 1993 : 25.

2. Wright, 1993 discute une objection à Horwich : recourir aux lois de l'univers n'explique pas la relation causale alléguée dans (iii). Mais si l'on admet (ii), et que « *p* » dit que *p*, on peut répondre à la question de savoir pourquoi les choses sont telles que « *p* » dit qu'elles sont en citant simplement le fait que *p*.

> Un énoncé peut être vrai sans que nous ayons jamais la possibilité de savoir qu'il l'est
> Tout énoncé susceptible d'être vrai a une négation, elle-même susceptible d'être vraie[1].

La liste n'est pas exhaustive. On pourrait ajouter : la vérité est stable (quand les phrases ne contiennent pas d'indexicaux ou de temps), un énoncé n'est pas plus ou moins vrai, mais vrai absolument, etc. Ces platitudes n'expriment rien de substantiel, ou de profondément métaphysique sur la vérité.

Nous pouvons alors franchir une étape supplémentaire dans notre raisonnement. Si nous acceptons le déflationnisme quant à la vérité, il n'y a plus aucun obstacle à accepter aussi l'idée que *connaître la signification d'une phrase c'est connaître ses conditions de vérité*. En effet, l'expression « conditions de vérité », même si elle ne fonctionne pas elle-même de façon décitationnelle comme « est vrai », peut être rendue équivalente à ce prédicat :

> (iv) « *p* » est vrai si les conditions de vérité de « *p* » sont réalisées
> (v) « *p* » est vrai si les conditions de vérité de « *p* » sont telles qu'on peut l'asserter

Savoir qu'une phrase est vraie, c'est savoir que ses conditions de vérité sont réalisées, savoir qu'on peut l'asserter, que ses conditions de vérité correspondent aux faits, qu'elles sont réalisées conformément à la façon dont la phrase dit que les choses sont, etc. En ce sens, la proposition qui fonde la théorie vériconditionnelle de la signification n'est elle-même qu'une platitude. Comme le dit McDowell :

> Il y a une connexion truistique entre la notion d'un contenu d'assertion et la notion familière de vérité (notion dont nous pouvons penser que la signification est précisément fixée par cette connexion) ; la connexion garantit, comme pure et simple platitude, qu'une spécification correcte de ce qui peut être asserté par l'énonciation assertorique d'une phrase ne peut être autre chose qu'une condition sous laquelle la phrase est vraie. Une proposition radicale à ce point serait la suivante : tant que les extrémités des théorèmes (que l'on peut penser

1. Wright, 1993 : 34.

> comme ayant la forme « *s*...*p* ») sont reliées de telle manière que, quoi que disent les théorèmes, nous pouvons les utiliser comme s'ils disaient quelque chose de la forme « *s* peut être utilisé pour asserter que *p* », cela n'a pas d'importance si nous écrivons, entre ces extrêmités, quelque chose qui produit une vérité dans ces circonstances ; notre platitude garantit que « est vrai si et seulement si » remplit ce rôle, et ceci nous fournit une cible plus maniable que celle qui ressortait précédemment. (Voir Davidson, 1967.) Mais même si nous ne faisons pas cette hypothèse radicale, mais en quelque sorte nous attaquons à la cible initiale, la platitude ne cesse pas d'assurer que nos théorèmes, s'ils sont acceptables, doivent spécifier ce que sont en fait les conditions de vérité des phrases mentionnées — même si, dans cette seconde option, ils ne se représentent pas explicitement comme jouant ce rôle (McDowell, 1981 : 229, cf. aussi 1976 : 46).

Nous pouvons poursuivre, sur le mode de la correspondance : comprendre une phrase, c'est savoir ce qu'elle énonce ; ce qu'elle énonce c'est qu'un certain état de choses est présent ; par conséquent quelqu'un qui comprend un énoncé saura quel état de choses est présent, et en ce sens connaîtra ses conditions de vérité, etc. Ces platitudes sont parfaitement inoffensives. Elles n'imputent à la signification aucune propriété « substantielle » ou profonde, d'après laquelle les conditions de vérité se tiendraient, pour ainsi dire, au-delà de la phrase, dans une sorte d'arrière-plan inaccessible, comme le voudrait le réalisme métaphysique externe, pas plus qu'elles n'identifient cette signification à une forme d'assertabilité ou de justification de la phrase manifestable dans des capacités pratiques. Nous pouvons appeler cette thèse *conception déflationniste de la signification* (DS).

Le point important est alors le suivant. Si nous comprenons la thèse (I) du réalisme selon Dummett au sens du déflationnisme quant à la vérité et à la signification, alors il n'y a plus de raison de considérer que (I), le schéma décitationnel pour la vérité et la conception vériconditionnelle de la signification implique (II), la transcendance de la vérité par rapport à la vérification, ni (IV) la vérité-correspondance, *sinon dans les sens truistiques de ces notions*. Par conséquent (I) n'implique pas la forme de réalisme métaphysique que l'argument antiréaliste standard prétend qu'elle implique. Elle n'implique pas non plus, du fait de la platitude concernant la relation de la vérité et de l'assertion, une forme de vérificationnisme. Mais (I) n'interdit pas à un réaliste métaphysique, s'il en a envie, d'interpréter la platitude concernant la correspondance en un sens « profond », pas plus

qu'elle n'interdit à un antiréaliste d'interpréter, s'il en a envie, la platitude concernant l'assertion en un sens également « profond ». Or il en est de même pour la relation entre (I) et (III), le principe de Bivalence : (I) ne nous engage pas à défendre ou à rejeter la logique classique bivalente. Etant donné que Dummett associe étroitement la thèse réaliste avec l'adoption de la Bivalence, et l'antiréalisme à un révisionnisme intuitionniste, cette connexion mérite d'être analysée à présent. Est-ce que (I) entraîne ou présuppose (III) ou est incompatible avec sa négation, c'est-à-dire avec le refus intuitionniste d'accepter la Bivalence (ou de la nier) pour certaines phrases ?

L'adoption du schéma (T) comme condition d'adéquation d'une théorie de la vérité et la conception vériconditionnelle présupposent-elles la Bivalence ? Non, car comme l'ont montré en particulier McDowell (1976), Evans (1976), et Tennant (1987), il est parfaitement possible d'adopter une théorie-T dont le métalangage obéit aux règles de la logique intuitionniste sans renoncer à la conception vériconditionnelle. Les définitions tarskiennes standard de la vérité utilisent une théorie de la démonstration classique, et par conséquent prouvent, pour chaque phrase du langage-objet, que cette phrase ou sa négation est vraie. Mais une théorie de la vérité dotée d'une logique intuitionniste ne prouverait rien de tel. Il ne s'ensuit pas que l'on ne puisse pas définir la vérité dans un métalangage intuitionniste, ni que l'on ne puisse considérer le langage-objet comme obéissant à une logique intuitionniste. La logique requise par une théorie-T est une logique minimale ou « de base », qui impose seulement une règle d'inférence par substitution sur les biconditionnels, et elle peut être formulée avec des constantes dotées des règles d'introduction et d'élimination usuelles. Elle n'impose aucune *interprétation* particulière, classique ou intuitionniste à ces opérateurs.[1] Ce n'est que si nous *ajoutons* le réquisit que nous devons adopter telle ou telle interprétation que nous nous trouverons commis à la Bivalence ou pas. Il devient donc parfaitement possible de défendre une position comme celle de Tennant 1987, qui défend la conception *antiréaliste* suivante :

1. Tennant, 1987 : p. 71-75, cf. également chap. 17 ; cf. aussi Engel, 1989, chap. XI, XII.

> (1) S est vrai *s* si *p* (où *p* est une traduction métalinguistique de S)
> (2) connaître la signification de S c'est savoir sous quelles conditions S est ou serait vraie
> Ces deux schémas survivent intacts à la critique antiréaliste. La bivalence est le point litigieux. Les deux schémas servent seulement à épingler la notion légitime de vérité. La bivalence est le revêtement classique qu'elle n'aurait jamais dû acquérir.
> La survie du second schéma, pour l'antiréaliste, est tautologique. Elle sert à caractériser ce que peut vouloir dire le terme « signification ». Elle impose le réquisit que cette notion soit sensible, de façon appropriée, à notre épistémologie de la compréhension.
> La survie du premier schéma est une marque de la correction de la notion antiréaliste de vérité. Comme nous l'avons vu, le réquisit tarskien d'adéquation sur une définition métalinguistique du prédicat de vérité peut être satisfait avec les moyens logiques les plus faibles dans le métalangage (Tennant, 1987 : 130).

D'après Tennant (§ 4.6), une conception antiréaliste de la signification peut parfaitement conserver les traits de compositionnalité et de molécularité qui appartiennent à une théorie-T. Elle peut aussi parfaitement conserver (1) et (2) (c'est-à-dire (I), la conception vériconditionnelle de la signification et le schéma décitationnel), comme conditions d'adéquation sur une théorie de la signification. Ce qui rendra cette théorie « antiréaliste », c'est le refus d'adopter la Bivalence et la logique classique. Mais ces derniers ne sont en rien inclus dans le paquetage (I). Le « réalisme » intervient selon Tennant quand on choisit d'interpréter la conception vériconditionnelle comme imposant l'idée que tout énoncé est vrai ou faux. Mais l'antiréaliste refuse ce choix :

> L'adoption de règles strictement classiques marque un saut unique, à partir d'une conception dont les règles sont justifiées à une autre. Ce n'est pas simplement un passage d'une justification d'un système limité à une justification disponible à un système plus large, qui serait effectué sur l'arrière-plan de la même conception de ce que nous faisons quand nous communiquons au moyen du langage et de la manière dont nous devrions en conséquence préserver, développer et représenter le contenu sémantique dans nos méthodes de raisonnement déductif. Il s'agit plutôt d'un écart radical — comme Dummett y a insisté — en direction d'une conception de la vérité transcendante par rapport à la vérification qui dépasse potentiellement nos capacités recognitionnelles. Le passage au classicisme enregistre un déplacement non justifié et philosophiquement suspect à

> une conception métaphysique. Il abandonne la conception du langage comme instrument social, et des capacités logiques et linguistiques comme fondées dans un comportement publiquement observable (Tennant, 1987 : 148).

La conception vériconditionnelle, et les platitudes qu'elle incorpore, est donc neutre par rapport au statut du principe de Bivalence. Cela confirme le fait que ce principe n'est pas la pierre de touche du réalisme, *si comme Dummett on voit dans (I) un principe « réaliste »*. (I) ne présuppose en rien la Bivalence. La Bivalence est bien la pierre de touche du réalisme, mais du réalisme *externe*, qui interprète d'une certaine façon les platitudes contenues (I). Mais si, comme on l'a vu, ces platitudes permettent aussi d'inclure l'idée d'une distinction entre verité et assertabilité garantie, et l'idée de correspondance, alors (II) et (IV) aussi sont des platitudes inoffensives.

Mais même si (II), la transcendance de la vérité n'est pas une platitude inoffensive, il n'est pas du tout évident qu'elle soit entraînée par, ni qu'elle entraîne, la Bivalence. Comme le dit Dummett (cf. § 4.2), l'idée qu'une phrase doit être vraie ou fausse nous incite à adopter l'idée d'une vérité dépassant toute reconnaissance. Mais Wright a contesté que ce lien soit nécessaire, au moins quand il s'agit d'adopter la Bivalence pour une certaine classe d'énoncés :

> Il est vrai que si quelqu'un accepte la Bivalence pour une classe d'énoncés dont nous ne pouvons pas dans tous les cas garantir les valeurs de vérité au moyen d'une décision, alors il s'engage au moins à soutenir que nous ne pouvons pas garantir que la vérité coïncide toujours avec la vérité décidable. Mais à moins qu'il accepte la transition de « nous ne pouvons pas garantir que P » à « c'est une possibilité que non-P » il ne s'est pas engagé à défendre la possibilité d'une vérité transcendante. Cette transition est intuitionnistiquement suspecte : de tout énoncé mathématique qui n'est pas effectivement décidable il serait intuitionnistiquement correct de dire que nous ne pouvons pas garantir l'existence de moyens de le vérifier ou de le falsifier, mais ce n'est pas, au vu de l'analyse intuitionniste de la négation, acceptable comme une possibilité intuitionniste qu'il n'y ait tout simplement pas de moyen de vérifier ou de falsifier l'énoncé en question. Il est par conséquent douteux que le fait d'endosser la Bivalence pour des énoncés autres qu'effectivement décidables soit par soi-même une admission de la possibilité d'une vérité transcendante (Wright, 1987 : 318).

Inversement, est-ce que l'idée que la vérité peut transcender notre pouvoir de vérification entraîne l'adoption ou le rejet de la Bivalence ? Selon McDowell, non : on peut très bien souscrire aux doutes antiréalistes quant à la Bivalence, et, en conséquence, proposer une théorie vériconditionnelle de la signification qui prenne la forme d'une théorie-T intuitionniste, et créditer des locuteurs dont cette théorie décrirait la compétence sémantique d'une connaissance des conditions de vérité des phrases, qu'elles soient réalisées ou pas, indépendamment du fait que nous disposons ou pas de moyen de les justifier[1]. Cela revient à dire que les platitudes incorporées dans la théorie vériconditionnelle de la signification sont parfaitement neutres par rapport au statut de (II), comme elles le sont vis-à-vis de (III) et de (IV). Nous pouvons alors nous demander ce qui reste du débat entre « réalisme » et « antiréalisme ». Il y a deux réactions possibles, si l'on accepte le déflationnisme quant à la vérité et quant à la signification. La première consiste à soutenir que les platitudes (I) sont en quelque sorte un terrain neutre, bien que nécessaire, à partir duquel on doit bâtir une théorie de la signification substantielle. C'est par exemple la position de Tennant 1987, qui adopte les platitudes, et développe une sémantique antiréaliste fondée sur une logique non classique (la logique intuitionnniste de la pertinence). Une position symétrique réaliste consisterait à admettre les platitudes, pour essayer de construire une théorie substantielle de la vérité-correspondance, peut-être en développant la notion de fait ou d'état de choses, ou en proposant une théorie causale de la référence. J'envisagerai une autre possibilité un peu plus loin[2]. Mais il y a une interprétation, plus radicale, ou plus littérale, du déflationnisme : c'est l'idée, déjà esquissée au § 5.1 *in fine*, qu'en dehors des platitudes portant sur la vérité et la signification, il n'y a « rien d'autre à dire » sur ces notions : pas de « théorie », métaphysique ou autre, à élaborer sur elles. Les platitudes, en quelque sorte, sont nécessaires *et* suffisantes. Ce sont des notions que nous devons laisser en paix. C'est pourquoi on peut appeler cette position *quiétiste*.

1. Mc Dowell, 1976 : 55.

2. Baldwin, 1991 : 33-35 fait allusion à cette possibilité. Forbes, 1987, va plus loin en essayant d'expliquer la notion de correspondance à partir de celle de fait. Cf. également Taylor, 1985. Il y a bien d'autres possibilités, comme par exemple Millikan, 1984. Je reviens à la théorie causale au chapitre suivant.

5.4. Le charme discret du quiétisme

Le quiétisme, en dehors de sa connotation religieuse, est en général la thèse, d'inspiration positiviste ou wittgensteinienne (« *Friede in den Gedanken* »), selon laquelle les discussions métaphysiques n'ont pas de sens[1]. Dans le contexte qui nous occupe, il est la thèse selon laquelle il n'est pas possible de formuler un débat significatif entre le réalisme et l'antiréalisme quant à la signification et la vérité, parce qu'on ne peut pas aller au-delà des platitudes déflationnistes. Cette thèse aussi a des liens étroits avec une certaine interprétation de Wittgenstein, que nous n'examinerons qu'au chapitre 7, ainsi qu'avec la conception « néo-pragmatiste » de philosophes comme Rorty (1979, 1981), qui entendent aller « au-delà » du débat entre réalisme et antiréalisme, que je mentionnerai au chapitre suivant.[2] Mais je la développerai ici sur d'autres bases. Johnston (1988) donne, à mon sens, une bonne caractérisation du quiétisme (qu'il appelle « minimalisme » quant à la signification) :

> (1) La signification n'a pas de nature cachée et substantielle que pourrait révéler une théorie de la signification. Tout ce que nous savons et avons besoin de savoir sur la signification en général est donné par une famille de platitudes [du type « 'S' signifie que p si en énonçant 'S' sur le mode assertorique, un locuteur asserterait que *p*]
> (2) Ces platitudes prises ensemble représentent le discours au sujet du potentiel d'une expression à pouvoir asserter, commander, demander, etc., diverses choses.
> (3) Par conséquent comprendre les significations des expressions n'est pas quelque chose qui les sous-tend et qui serait la base causale explicatrice de l'aptitude à utiliser les expressions pour asserter, commander, demander, etc., diverses choses. Il est plutôt constitué par cette aptitude.
> (4) Par conséquent une théorie de la signification pourrait au mieux être un énoncé de propositions dont la connaissance nous permettrait d'acquérir l'aptitude pratique. Mais à cet égard un manuel de traduction servirait aussi bien. Par conséquent l'intérêt d'une théorie de la signification est minimal et il ne fait pas de doute qu'aucun problème intéressant quant à l'objectivité et au réalisme ou quant à la relation de l'esprit et de la réalité ne peut être posé en considérant des questions portant sur la forme d'une théorie de la signification.

1. Il ne s'agit pas évidemment d'assimiler ces positions, mais seulement d'indiquer leur « air de famille ». Je reviens sur la question du quiétisme de Wittgenstein au § 7.5. A ma connaissance, le terme est dû à Blackburn, 1985, 146, qui parle aussi de *dismissive neutralism*.

2. Cf. chap. 6, note 32, en fait le quiétisme, comme on l'aura noté, a aussi des affinités avec la position de Davidson, que j'examinerai au chap. 6, mais aussi d'auteurs « nihilistes » tels que Kripke (1981) et Schiffer (1987), que j'examinerai au chapitre 7.

Les propositions (1), (2) et (3) correspondent en gros à ce que nous avons appelé la conception déflationniste de la signification. (3) est la conclusion proprement quiétiste. La transition de l'une à l'autre est facile. Il semble en fait que le déflationnisme soit déjà un quiétisme ; mais il est tout à fait important pour la suite de notre argumentation de les distinguer. Comme nous l'avons vu, on peut, comme Tennant, admettre DV et DS, et refuser la conclusion quiétiste. On peut être frappé par le fait que la position décrite par Johnston est très proche de ce que Dummett appelle une théorie « modeste » de la signification. Et de fait le principal représentant du quiétisme, McDowell, a défini sa propre position comme « une défense de la modestie » contre Dummett[1]. Mais il l'appelle également « réalisme trivial » ou « simple ».

Selon McDowell, il n'y a rien de plus dans la vérité et la signification que les platitudes déjà mentionnées. Elles n'entraînent ni la Bivalence, ni la transcendance du sens par rapport à l'usage, ni la possession par les locuteurs de pouvoirs mystérieux de recognition d'états de choses indétectables. Ces truismes ne nous engagent pas à souscrire au réalisme externe. Mais ils ne nous engagent pas non plus à souscrire à une forme de vérificationnisme comme celle de Dummett. Comme on l'a vu, McDowell accuse Dummett de donner une interprétation beaucoup trop restrictive du slogan wittgensteinien « le sens c'est l'usage » et de souscrire à un béhaviorisme implicite. Interpréter correctement ce slogan, c'est admettre que la position que Dummett décrit comme propre à une théorie substantielle, d'après laquelle nous devrions pouvoir déterminer tout contenu de signification « de l'extérieur », du point de vue de quelqu'un qui ne comprendrait *pas* un langage et qui voudrait l'apprendre, est intenable. Selon cette conception,

1. Mc Dowell, 1987 ; cf. aussi 1981, 1984.

> Toute activité humaine intelligible peut être décrite de manière à révéler son but ou son sens à partir de la perspective d'un exil cosmique — c'est-à-dire une perspective qui n'est en aucune manière colorée ou affectée par l'implication de son occupant dans une forme de vie ; car la capacité de la description à rendre l'activité compréhensible ne doit pas dépendre d'une telle implication (McDowell, 1981 : 237).

C'est au nom de cette conception que Dummett s'oppose à l'idée, qu'il tient comme propre à une théorie modeste de la signification, selon laquelle une telle théorie doit déterminer une signification qui serait déjà comprise par les locuteurs dont on veut caractériser la compréhension. Mais pour McDowell, ce point de vue est inévitable, et nous devons rejeter la perspective de « l'exil cosmique ». Comprendre un langage, c'est déjà, et simultanément, être impliqué dans une certaine « forme de vie », par laquelle les mots que nous employons sont doués d'un sens que nous saisissons immédiatement. Une théorie de la signification, qui prend la forme d'une théorie homophonique de la vérité, ne fait pas autre chose que spécifier ces significations en tant qu'elles sont déjà comprises et manifestées dans un langage public. Cela implique que nous ne considérions pas la compréhension d'un langage comme une sorte d'hypothèse, que nous imposerions de l'extérieur du locuteur, sur des propriétés de son comportement, ou comme une certaine capacité inférentielle médiate de recognition, caractérisable psychologiquement, mais comme une forme de perception, donnée de manière immédiate par l'implication des locuteurs dans la pratique linguistique. C'est en ce sens, non psychologique, que « le sens c'est l'usage », et cela entraîne, aux yeux de McDowell, une répudiation de l'idée qu'une théorie de la signification puisse être autre que modeste, autre que caractérisant la signification *de l'intérieur* de la pratique linguistique elle-même. Cela entraîne également que le mental ne soit pas considéré comme extérieur au langage, mais directement présent en lui, à l'opposé même de la conception du langage comme code :

> Le projet de Dummett était d'éviter le psychologisme tout en reconnaissant la connexion du langage avec l'esprit. Or l'implication du langage dans l'esprit dans le discours doué de sens est précisément la possession d'un contenu par les énonciations. La voie moyenne, donc, consiste à abandonner l'idée qu'une théorie de la signification devrait chercher à caractériser les énonciations de manière

> à les représenter comme signifiantes, autrement qu'en leur attribuant précisément des contenus. Une fois que nous avons éloigné l'implication supposée de la modestie dans la conception du langage comme code, nous pouvons voir précisément que cela ne revient pas à retomber dans le psychologisme. Le rejet du psychologisme est la thèse que les sens des énoncés ne sont pas cachés derrière eux, mais se tiennent visibles aux yeux de tous : c'est-à-dire qu'être un locuteur d'un langage, c'est être capable de mettre ses pensées dans ses mots, là où les autres peuvent les entendre. La grande beauté de la conception modeste de la signification « homophonique » est qu'elle franchit la distance destinée à rendre l'idée non problématique, en montrant que nous n'avons pas besoin de penser qu'elle fasse plus que cela : la pensée, par exemple, que certaines nappes sont carrées peut être entendue dans les mots « Certaines nappes sont carrées », par des gens qui sont capables de mettre leurs esprits, s'ils en ont l'occasion, dans ces mots (McDowell, 1987 : 30).

Si on admet que comprendre un langage n'est autre chose que cette forme de perception immédiate du sens découlant de notre immersion dans une pratique publique (« la signification comme physiognomie », Wittgenstein (1951, 568)), alors il n'y a plus aucune raison de chercher à *rendre compte* de cette capacité, ou de chercher à l'*expliquer*, c'est-à-dire de fonder une analyse de la signification qui en ferait le produit d'une *théorie* implicite possédée par les locuteurs. On mécomprend complètement la phénoménologie de la compréhension du langage quand on cherche à analyser cette compréhension comme une forme de *connaissance fondée*. La seule élucidation que nous puissions donner de la signification est une forme de *description* de notre pratique.

Le « réalisme trivial » n'est donc un réalisme qu'en un sens prudhommesque. Il n'est aussi, si l'on peut dire, un antiréaliste qu'en un sens pickwickien. La conclusion s'impose : ce « réalisme » est dans la même position vis-à-vis du réalisme classique ou externe et vis-à-vis de l'antiréalisme dummettien que l'idéalisme transcendantal kantien par rapport au réalisme transcendantal et à l'idéalisme empirique. Mais McDowell préfère plutôt comparer sa position à celle du point de vue « grammatical » de Wittgenstein :

> La volte-face entre la première et la seconde philosophie de Wittgenstein est en partie un remplacement du réalisme par l'antiréalisme à son niveau transcendantal...

> Le réaliste transcendantal soutient que du point de vue de l'exil cosmique on pourrait être capable de discerner des relations entre notre langage et un monde conçu de manière réaliste. Les antiréalistes se récrient avec raison, mais de manières différentes. L'antiréaliste quant à la signification se récrie en donnant une image différente de ce à quoi les choses ressembleraient à partir de cette perspective ; mais la bonne solution est de détourner sa face de l'idée d'un exil cosmique (McDowell, 1981 : 248)[1].

5.5. Le réalisme minimal

Il y a dans le déflationnisme, le quiétisme, ou les diverses formes de minimalisme que nous venons d'examiner une sorte de délectation à décevoir les attentes que les philosophes peuvent placer dans le projet d'une « théorie de la signification » du type de celui qui nous occupe depuis le début de ce livre. A quoi bon cette enquête, si nous devons nous retrouver dans la situation du dogmatisme face à la dialectique transcendantale, ou du métaphysicien face à la thérapeutique wittgensteinienne ? Nous aurions poursuivi la signification jusque dans une impasse. Mais à ce quiétisme on doit répondre : nos attentes doivent-elles être *nécessairement* déçues ? Tout d'abord, malgré l'insistance du quiétiste à dire que ce sont des platitudes anodines, il est très difficile de résister à l'intuition qu'il doit y avoir *plus* dans la vérité et dans la signification. Le quiétiste ne dit pas seulement qu'on doit s'en tenir aux platitudes. Sa position semble menacer l'*objectivité* même de la vérité et de la signification (et en ce sens se rapprocher d'un certain *nihilisme* quant à ces notions)[2]. Or nous ne sommes pas prêts à renoncer à l'idée que la vérité est indépendante de sa justification, qu'elle consiste dans une certaine correspondance avec une réalité autonome, que l'on doit pouvoir dire plus, pour expliquer le sens de « la neige est blanche », que c'est ce que « la neige est blanche » dit et que cette phrase est vraie quand ses conditions de vérité sont réalisées[3].

1. Cf. également Bouveresse, 1981 : qui cite approbativement ce passage. McDowell associe également sa position à une forme de théorie romantique du langage comme celle de Herder. J'ai examiné ce point dans Engel 1990.
2. Je reviens sur ce point au chap. 7.
3. Cf. la réaction de Dummett, 1987, citée au § 7.5, *in fine*.

Et nous ne voulons pas non plus renoncer à l'idée qu'il existe un lien entre la vérité et la connaissance de la vérité plus « épais » que la trivialité selon laquelle dire que *p* est vrai c'est l'asserter. Mais le quiétisme nous répondra, suavement, que sa théorie rend compte de ces intuitions. Que sont-elles de plus alors que des *intuitions* ?

Il y a deux manières d'expliciter ces intuitions. On peut, d'abord, rejeter la prémisse déflationniste du quiétiste, en niant que les platitudes *soient* des platitudes. On peut, ensuite, concéder au quiétiste sa prémisse déflationniste, mais refuser de dire que les platitudes sont *seulement* des platitudes. Dans les deux cas, nous devrons nous engager à chercher une conception qui aille *au-delà* des platitudes. Mais alors ne nous retrouvons-nous pas tout simplement face à notre débat initial entre réalisme et antiréalisme ? La voie est étroite. Mais on peut essayer de définir une position qui admette :

(a) l'objectivité de la vérité mais pas sa réduction à l'assertabilité garantie (sans quoi nous revenons au vérificationnisme),
(b) que la vérité est soumise à des contraintes épistémiques (sans quoi nous revenons au réalisme externe),
(c) l'objectivité de la signification (sans quoi ce n'est plus la peine de parler d'une « théorie » de la signification),
(d) mais aussi des contraintes épistémiques sur cette notion (sans quoi nous divorçons le sens de l'usage),
(e) les platitudes déflationnistes (sans quoi nous ne pouvons plus reconnaître le lien intrinsèque entre vérité et assertion, vérité et signification).

Une telle position est déflationniste, mais elle n'est pas quiétiste, car elle admet qu'il y a quelque chose d'intéressant à dire sur la vérité, la signification, et sur le sens « métaphysique » de ces notions. Je l'appellerai *réalisme minimal*. En fait il existe une famille de positions de ce type. Je discuterai ici surtout les conditions (a), (b) et (e) qui portent sur la vérité, et j'examinerai les conditions portant sur la signification dans les chapitres suivants.

Les lecteurs de Putnam (1981, 1990) pourraient reconnaître dans le portrait-robot que je viens d'esquisser ce qu'il appelle *réalisme interne*. Putnam rejette le réalisme externe ou métaphysique, mais admet que celui-ci a raison de divorcer la vérité de l'assertabilité garantie. Il refuse donc le

vérificationnisme, mais entend maintenir un lien entre vérité et assertabilité. Sa solution consiste à dire que la vérité est une *idéalisation* de l'assertabilité garantie ou de l'acceptabilité rationnelle. La vérité est ce qui *serait* justifié dans des circonstances épistémiques idéales. On trouve une idée voisine chez Peirce, qui définit la vérité comme ce qui correspondrait à la réalité « à la limite de l'enquête scientifique », et chez Popper[1]. Elle ressemble aussi à ce que nous avons appelé au chapitre précédent (§ 4.1) « vérificationnisme idéal ». L'objection que l'on adresse en général à cette position est qu'on ne voit pas bien ce que pourraient être ces conditions « idéales » de vérification. Si nous pouvons parler de conditions d'assertion garantie ou de vérification, ce sont des conditions que nous connaissons et pouvons formuler. Des conditions « limites » semblent tout aussi informulables et inaccessibles que la correspondance de la réalité avec nos énoncés postulée par le réalisme externe (*a fortiori* si l'on associe la vérification idéale à une forme de *correspondance*). Comme le dirait Russell, nous ne pouvons les atteindre, au moins pour des raisons « médicales ». Et comme on l'a vu, des conditions idéales d'assertabilité violent le réquisit de manifestation dummettien du sens : elles seraient tout aussi peu manifestables et compréhensibles que les conditions de vérité « indétectables » du réaliste sémantique. A cela Putnam répond que nous pouvons quand même « approximer » les conditions idéales « à un degré élevé ». Mais comment formuler ce degré et cette approximation ? Ces difficultés sont bien connues. Mais il y a une difficulté supplémentaire, qui a été formulée par Wright (1992 : 39-44). Le réalisme interne, s'il doit satisfaire à nos conditions (a)-(e), doit admettre les platitudes déflationnistes quant à la vérité. Il doit par conséquent soutenir que sa définition du prédicat de vérité :

S est vrai ssi S serait justifié dans des conditions épistémiques idéales

possède les propriétés décitationnelles usuelles. Or Putnam impose ce qu'il appelle un réquisit de *convergence* à sa définition : qu'il n'y ait pas d'énoncé tel que cet énoncé et sa négation soient assertables dans des

1. Pour une comparaison des vues de Peirce avec celle de Putnam, C. Tiercelin, 1990. Mais Putnam (1990) nie que sa conception soit identique à celle de Peirce.

conditions épistémiques idéales. Autrement dit, les enquêteurs à la limite devraient converger sur l'assertabilité ou non d'un énoncé. Mais cela ne veut pas dire que dans les conditions idéales, pour tout énoncé, cet énoncé *ou* sa négation devraient être justifiés, ou que tout énoncé devrait être décidable dans ces conditions idéales. Bref on ne peut affirmer la Bivalence dans les conditions idéales. Mais si les platitudes sont correctes, alors la négation de l'équivalence usuelle

« Il n'est pas le cas que S » est vrai ssi il n'est pas le cas que « S » est vrai.

Mais si « vrai » veut dire « assertable idéalement » et qu'on admet qu'on pourrait se trouver dans une situation épistémique dans laquelle ni S ni sa négation ne sont assertables, alors il faudrait rejeter la négation de l'équivalence, puisqu'on ne pourrait pas dire dans les conditions idéales qu'il n'est pas le cas que « P » est assertable. On ne voit pas alors comment le réalisme interne peut rester neutre quant à l'affirmation ou la négation d'un énoncé, et suspendre sa croyance quant à la Bivalence. Nous nous retrouvons dans une situation familière : ou bien nous maintenons l'identité de la vérité et de l'assertabilité garantie (EV du § 5.2), et, en cas, il nous faut admettre la conception intuitionniste de la vérité et rejeter l'idée d'une indépendance de la vérité par rapport à nos justifications que le réalisme interne veut quand même maintenir, ou bien nous rejetons cette identité, mais alors comment définir la vérité comme assertabilité ou acceptabilité rationnelle idéale ? Le réalisme interne est une thèse très instable.

Si la position que nous essayons de formuler entend, tout en admettant les intuitions réalistes de base, accepter ce qui constitue, finalement, l'idée fondamentale de l'antiréalisme, à savoir qu'il y a un *lien*, sinon d'identité du moins très étroit, entre vérité, assertion et assertabilité, alors il est manifeste que nous devons chercher à définir plus précisément les conditions d'assertabilité, comme l'illustrent les difficultés du réalisme interne. Après tout, si nous admettons qu'il y a des contraintes épistémiques sur la notion de vérité et sur la notion de signification, il faudra bien que nous introduisions des considérations épistémologiques dans nos théories de la signification et de la vérité.

On a souvent remarqué que Dummett, au moins dans ses premiers écrits, tend à transposer la notion de « condition d'assertion » du contexte des mathématiques (intuitionnistes) à celui du langage naturel. Dummett est cependant bien conscient des difficultés de cette transposition :

> La différence principale entre le langage des mathématiques et notre langage en général tient au fait que, dans le premier, la propriété de décidabilité est stable. Un énoncé disant qu'un certain nombre particulièrement grand est premier peut en principe être décidé, et il est par conséquent légitime d'asserter la disjonction de cet énoncé et de sa négation, ou tout autre énoncé qui peut être démontré suivre de l'énoncé et de sa négation, car à tout moment où nous le voudrions nous pourrions, au moins en théorie, déterminer l'énoncé comme vrai ou faux. Mais la décidabilité d'un énoncé empirique n'est pas de la même manière un trait constant : si nous considérons l'énoncé « Il y a maintenant ou bien un nombre pair ou bien impair de canards sur la mare » comme assertable sur la base du fait que nous pourrions, si nous le voulons, déterminer l'un ou l'autre membre de la disjonction comme vrai, nous ne pouvons pas offrir le même fondement pour l'assertion de « Il y avait soit un nombre pair soit un nombre impair d'oies sur le Capitole » et si, cependant, nous voulons considérer ce dernier énoncé comme assertable, alors ou bien on doit expliquer cette assertion comme énonçant une relation plus faible que la vérifiabilité de principe, ou bien la spécification de ce qui compte comme vérifiant un énoncé complexe ne peut pas être donnée de manière uniforme en termes de ce qui compte comme vérifiant les constituants (Dummett, 1976 : 113-114).

Il y a ici deux difficultés. La première est que la vérification d'un énoncé empirique peut ne pas être définitive : elle peut être *annulable (defeasible)* ou révisable, selon l'information dont nous disposons. C'est pourquoi la vérifiabilité ou l'assertabilité doit être une notion plus faible que celle de vérifiabilité en principe ou de démonstrativité. La seconde est que l'on voit mal, dans le cas non mathématique, comment maintenir le principe de compositionnalité pour des énoncés comme les énoncés disjonctifs si l'on remplace la vérité par la vérification. Il est clair en tout cas que Dummett est beaucoup moins confiant dans l'idée que la notion d'assertabilité puisse être assimilée à celle de justification complète et fondée que le soutiennent les critiques qui, comme McDowell, lui attribuent une sorte de fondationnalisme, et loin également du vérificationnisme viennois qui supposait que l'on puisse définir strictement les conditions de vérification

des énoncés. Les tentatives les plus élaborées de formulation de la notion d'assertabilité me paraissent être celles de Wright (1987, 1992) et de Peacocke (1986, 1987, 1992).

Wright (1987) a cherché à élaborer, dans un cadre qu'il voulait au départ antiréaliste, une notion plus faible d'assertabilité. Il s'est d'abord inspiré de la notion wittgensteinienne de « critère ». Un critère est une certaine *raison* pour asserter un énoncé, sans être un *fondement* certain ; les critères sont annulables et pluriels (il peut y avoir plusieurs critères pour une assertion), et les critères au sens wittgensteinien sont en principe manifestables et publics[1]. Mais cette notion ne peut pas se substituer à celle de condition de vérité, parce qu'elle manque de la *stabilité* qui est supposée appartenir à cette dernière — les critères pour un type d'énoncé peuvent varier selon les contextes — et parce qu'elle s'intègre mal dans le projet d'une théorie systématique de la signification[2]. Ces difficultés illustrent le fait que nous ne devons pas affaiblir la notion d'assertabilité jusqu'à la rendre indexicale, en faisant varier les conditions d'assertion selon le temps ou les sujets. Mais nous ne devons pas non plus la rendre trop forte, comme le montrent les problèmes inhérents à la notion d'assertabilité idéale. Selon la version putnamienne, un énoncé assertable dans des conditions épistémiquement idéales devrait à la fois n'être pas annulable (car autrement la vérité ne serait pas une propriété stable de nos énoncés) ni révisable à la lumière d'informations supplémentaires (sans quoi il ne serait pas « idéal »). C'est pourquoi il faut tenir les circonstances d'assertion comme idéales, ce qui voudrait dire que nous disposons (à la limite de l'enquête) de *toutes* les informations empiriques disponibles qui nous permettent de garantir à ces énoncés ces propriétés. Mais c'est évidemment demander, ou idéaliser, bien trop. Wright propose que nous considérions un énoncé comme assertable quand il est justifié par un certain état contrôlable d'informations et qu'il *demeure* tel *après* un élargissement éventuel des informations[3]. Il appelle cette propriété *surassertibilité*

1. Wright, « Truth conditions and criteria » (1976), « Anti-realist Semantics : the Role of Criteria » (1978), « Second Thoughts on Criteria » (1982), *in* Wright, 1987.

2. Comme n'ont pas manqué de le faire remarquer les wittgensteiniens « orthodoxes ». Cf chap. 7.

3. Wright, 1987 : 295-309, 1993 : 47-70. Les lecteurs de Popper auront reconnu une notion proche de celle de « corroboration » : acceptabilité d'un énoncé qui a résisté aux tentatives de réfutations. Mais, à ma connaissance, Wright ne fait pas le rapprochement.

[*superassertibility*]. Un énoncé est *surassertable* s'il est à la fois possible, dans des circonstances favorables, de parvenir à une certaine quantité d'information qui garantisse son assertion, et impossible, quelles que soient les circonstances, d'élargir cette quantité d'information de façon à ce que l'énoncé devienne révisable. Cette définition est évidemment vague, tant qu'on ne précise pas ce que peuvent être les circonstances de révision, ni ce que peut vouloir dire un « élargissement » de notre quantité d'information. Mais si l'on peut donner un sens à cette notion, elle remplit certains de nos réquisits : elle suppose une stabilité des énoncés surassertables, elle implique une idéalisation suffisante pour que les conditions d'assertions ne soient pas variables, mais cette idéalisation n'est pas trop forte ; elle ne présuppose pas que nous ayons une conception générale de ce que serait un énoncé qui ne puisse jamais être révisé. Le problème délicat, comme pour l'acceptabilité idéale, est de savoir si « est surassertable » peut fonctionner *comme* un prédicat de vérité, c'est-à-dire respecter les platitudes décitationnelles, sans *être* pour autant la vérité. Comme précédemment, nous pouvons nous demander si la version correspondante du schéma décitationnel

> « S » est surassertable ssi S

est correcte, et si on peut en inférer l'équivalence pour la négation

> « Il n'est pas le cas que S » est surassertable ssi il n'est pas vrai que « S » est surassertable.

Comme pour l'assertabilité dans les conditions épistémiques idéales, cela ne revient-il pas à confondre une certaine sorte de garantie (très) forte pour l'assertion de S avec la vérité de S, et l'une ne peut-elle pas intervenir sans que l'autre intervienne ? La réponse de Wright est que même si ce n'est pas assuré, nous pouvons néanmoins tenir la surassertabilité comme un *modèle* possible pour la vérité[1].

Mais ici le déflationniste (et sans doute le quiétiste) pourront demander : si un prédicat-substitut de vérité, comme la surassertabilité, doit

1. Wright, 1992 : 57. Mais il faut pour cela ajouter des suppositions qui ne font pas partie des platitudes. Je ne détaille pas sa discussion très complexe ici (50-56).

satisfaire aux platitudes, pourquoi ne pas tout simplement s'en tenir tout simplement au prédicat de vérité lui-même, et aux platitudes ? La réponse de Wright est que bien que le déflationnisme soit correct, il n'est pas *totalement* correct, parce qu'il y a *quelque chose de plus* dans le prédicat usuel de vérité que ce qu'exprime le schéma décitationnel. Il est faux, selon lui, que « vrai » fonctionne *seulement* comme un prédicat décitationnel et un dispositif d'assertion. Pour que l'équivalence entre « *p* » *est vrai* et l'assertion que *p* vaille, il faut qu'il existe dans le langage une *norme* de l'assertion correcte, c'est-à-dire une distinction entre les usages corrects et incorrects de phrases dont le contenu est assertorique. Cela se rattache étroitement à ce que nous avons remarqué à la fin du chapitre 3 : « est vrai » ne fonctionne pas seulement dans le schéma décitationnel comme prédicat de vérité ; il exprime également une norme réglant la pratique de l'assertion, une propriété qui nous *prescrit* ce que nous devons faire pour nous conformer à une pratique assertorique. Le prédicat de vérité exprime donc plus que les propriétés formelles auxquelles le réduit le déflationnisme. Il exprime une propriété normative, ce que nous avons appelé à la suite de Wiggins, la *Vérité* comme norme de l'assertion, que nous avons énoncé à la suite de Wiggins comme une « marque de vérité » :

> (i) Le désir des locuteurs de donner leur assentiment à une phrase doit, en général, être une indication de leur croyance qu'elle est T ; T doit être une propriété vers laquelle tendent normalement les assertions.

Mais alors on doit dire qu'il y a plus dans les platitudes que ce que dit le déflationnisme : la vérité est bien, en ce sens, une propriété « substantielle ». Wright appelle la conception des platitudes déflationnistes et de la norme d'assertion *vérité minimale*[1]. Quel est alors le lien entre cette norme de *Vérité* et la notion de surassertabilité proposée par Wright ? Ce n'est pas un lien d'identité : dire qu'il y a une norme de la pratique assertorique incorporée dans le prédicat « est vrai » n'implique pas une réduction de ce prédicat à l'assertabilité — ce qui aurait les conséquences désastreuses que nous avons déjà rencontrées avec EV au § 5.2 (voie(a) p. 193 —

1. Wright, 1992 : 15-32. Il distingue en ce sens son « minimalisme » de celui du déflationnisme (par exemple Horwich, 1990).

ni à la surassertabilité. Mais cette dernière est une *explication* des platitudes et de la propriété normative de vérité. Les platitudes énoncent les conditions auxquelles doit répondre tout prédicat de vérité pour être un prédicat de vérité. Mais nous pouvons interpréter, dans un second temps, ce prédicat comme étant la surassertabilité.

Les platitudes n'incluent cependant pas seulement la norme d'assertion (i). Comme nous l'avons vu, elles incluent aussi l'idée de correspondence. Wiggins a proposé d'ajouter à (i) quatre autres « marques de la vérité » :

(ii) le fait qu'une phrase soit T devrait, dans des circonstances favorables, conduire à une tendance des opinions concernant la correction d'un assentiment à converger ;
(iii) le fait qu'une phrase soit T ou pas, devrait être indépendante de tout moyen particulier qu'aurait un locuteur de reconnaître ce fait ;
(iv) quand une phrase est T, il devrait y avoir quelque chose en vertu de quoi elle est T ;
(v) quand une paire de phrases est T, leur conjonction l'est aussi (1980 : 209).

Le programme de Wright consiste à développer l'analyse de ces marques, en proposant notamment des critères de convergence des opinions[1]. Il propose notamment l'idée que nos jugements, dans des conditions épistémiques idéales, permettent de déterminer l'extension d'un prédicat ou d'un concept. Si nous prenons un certain concept, et pouvons enregister la meilleure manière selon laquelle les sujets peuvent déterminer l'extension de ce concept, il deviendra *a priori* vrai que ce concept s'applique. « La vérité, pour les jugements qui passent ce test, est constitutivement ce que nous jugeons être vrai quand nous opérons dans des conditions cognitivement idéales. »[2] La question qui se pose est évidemment celle de savoir comment nous pouvons établir ces conditions. Je reviendrai sur cette notion de convergence dans les chapitres suivants. Wright refuse, cependant, l'idée qu'on pourrait développer de manière cohérente la notion de correspondance au-delà de la platitude selon laquelle les énoncés vrais

1. Wright, 1987 ; 1992, chap .3.
2. Wright, 1989 : 246. L'exemple favori de Wright est celui des concepts de qualités secondes. Ils sont à la fois « dépendants de nos réponses » ou de nos jugements (au sens de Johnston, 1991) et objectifs si nous pouvons déterminer un accord sur les jugements que nous portons.

le sont « en vertu » de faits. C'est ce qui fait de son minimalisme un *anti-réalisme*, au sens où pour lui la vérité ne peut être qu'une notion liée à celle de connaissance et à des contraintes épistémiques.

On peut pourtant se demander ce qui différencie ce minimalisme de la position correspondante du minimalisme *réaliste* tel que nous l'avons défini par les conditions (a)-(e) ci-dessus. Qu'est-ce qui distingue le prédicat « vrai » du prédicat « surassertable » ? Il y a ici, fait remarquer Wright, une analogie avec le dilemme de l'*Eutyphron* de Platon, où l'on se demande si c'est *parce que certains actes sont pieux qu'ils sont aimés des dieux* ou si c'est *parce qu'ils sont aimés des dieux que certains actes sont pieux.* Nous pouvons nous demander de même si

c'est parce que certains énoncés sont vrais qu'ils sont surassertables

ou bien si

c'est parce que certains énoncés sont surassertables qu'ils sont vrais[1].

Le réaliste, au sens dummettien, admet la première proposition. Il identifie la source de l'assertabilité comme étant une vérité indépendante qui l'*explique*. L'antiréaliste, dans la réinterprétation de Wright, admet la seconde. Mais il n'identifie pas pour autant vérité et assertabilité. Selon Wright, on peut néanmoins admettre la première position ; cela dépend du type de discours concerné, selon qu'il s'agit des propriétés morales, des qualités premières et secondes, ou des énoncés mathématiques. De ce point de vue, le débat réalisme/ antiréalisme n'est pas homogène.

Il existe une autre manière d'établir un lien entre conditions d'assertion et conditions de vérité d'un énoncé sans réduire la vérité ni à l'assertabilité ni à une correspondence avec des faits transcendants. C'est celle que propose Peacocke (1986). Peacocke propose d'appeler « conditions d'acceptation » d'un certain contenu (sémantique ou de pensée) un certain ensemble de conditions qui nous font accepter rationnellement ce contenu. Mais Peacocke ne considère pas ces conditions d'acceptation rationnelles comme des conditions idéales au sens de Putnam. Il les considère comme des conditions normatives gouvernant *notre* acception effective

1. Wright, 1992 : 79-82 ; 108-139.

de contenus de pensée variés. Ce sont par exemple pour un contenu de perception dans lequel un sujet juge qu'un certain objet (un bloc) est cubique un « engagement canonique » pour un sujet tel que :

> (c) Pour toute position à partir de laquelle il percevrait ce bloc en *t* dans des conditions externes normales quand ses mécanismes perceptuels fonctionnent minimalement, il ferait l'expérience du bloc dans cette position relative comme étant cubique, ou qu'un objet cubique serait perçu dans cette position relative (Peacocke, 1986 : 15-16).

qui spécifie certaines conditions normales de perception. Et Peacocke propose l'idée que les conditions d'acceptation en ce sens *déterminent* les conditions de vérité du contenu en question. En d'autres termes, si une (ou des) condition(s) comme (c) sont remplies, alors le contenu « Ce bloc est cubique » est vrai. Inversement, s'il est vrai, ses conditions d'acceptation sont remplies. La stratégie de Peacocke consiste à étendre cette proposition à différentes classes de contenus. Par exemple, pour des énoncés contenant des termes logiques comme « et » ou « ou », on dira que les conditions d'acceptation sont fournies par les règles d'introduction et d'élimination usuelles de ces connecteurs, et que ces règles déterminent les conditions de vérité des énoncés conjonctifs ou disjonctifs correspondants[1]. La proposition de Peacocke ne revient *pas* à identifier les conditions de vérité aux conditions d'assertion (ou, dans le cas logique, de démonstration), mais elle établit un lien étroit entre les secondes et les premières. Le cas le plus intéressant est celui où les conditions de vérité des énoncés semblent engager le locuteur à saisir des conditions de vérité par nature transcendantes au sens dummettien, comme les énoncés impliquant une quantification universelle sur des lieux ou des temps *par définition* inaccessibles. Dans ce cas, Peacocke propose que nous considérions les conditions d'acceptation comme étant celles où le sujet juge, par exemple dans un certain lieu, que *tous les F sont G*, et où il suppose que dans tous les lieux accessibles spatialement à partir de ce lieu, il est vrai que

1. Cf. en particulier Peacocke, 1987. Cette thèse a des implications très importantes pour la question du sens des constantes logiques, que j'ai examinées dans Engel, 1989 ; dans la version de 1991 de ce livre, j'ai appelé « réalisme minimal » la position de Peacocke, parce qu'elle n'entraîne pas un vérificationnisme ni une réduction des conditions de vérité à des conditions de démonstration.

tous les F sont G. En d'autres termes, le sujet *étend* ou *projette* une quantification sur un domaine restreint à un domaine généralisé[1]. Nous ne supposons donc pas que le sujet a accès à des conditions idéales ou transcendantes, mais que ce qu'il juge comme la norme pour le domaine de quantification détermine les conditions de vérité de son jugement. Cette stratégie est très proche de celles que nous venons d'examiner. Mais Peacocke, à la différence de Wright, se considère comme un réaliste : il admet qu'une phrase puisse être vraie sans jamais pouvoir être vérifiée[2]. Mais il soutient qu'il doit y avoir des contraintes épistémiques sur la signification, précisément celles qui sont constituées par conditions d'acceptation normatives.

Toutes les conceptions que nous avons examinées ici montrent qu'un réalisme minimal au sens défini ci-dessus est au moins une position cohérente. Il s'ensuit que le débat entre réalisme et antiréalisme, tel que le concevait Dummett, prend une forme assez différente de sa forme initiale. Nous n'avons vu aucune raison de rejeter la conception vériconditionnelle de la signification (i) ni le schéma décitationnel. Ce que nous avons appelé les platitudes déflationnistes forme donc le noyau commun au réalisme et à l'antiréalisme. Contrairement au quiétiste, nous voulons aller au-delà de ces platitudes. Nous pouvons alors formuler une conception qui reste réaliste au sens où elle rejette une assimilation de la vérité à l'assertabilité, mais qui souscrit néanmoins à des contraintes antiréalistes, comme celles de surassertabilité ou d'acceptation rationnelle. C'est également une position qui souscrit à une conception épistémique de la signification (une théorie de la signification est une théorie de la compréhension) sans souscrire à une conception épistémique de la vérité. Si cette position est cohérente, alors le débat entre réalisme et antiréalisme perd une partie de sa substance sur ce point au moins : il n'y a plus de connexion évidente ou nécessaire entre les questions *sémantiques*, ou relevant de la théorie de la signification, et les questions *ontologiques*, relevant de la constitution métaphysique de la réalité. Une partie de l'attrait de l'argument antiréaliste standard venait de ce lien entre sémantique et ontologie, mais nous l'avons perdu. Nous pouvons à présent envisager une

1. Peacocke, 1986, chap. 3. et chap. 5.

2. Peacocke, 1987. Dans 1986a, Peacocke appelle sa position « manifestationnisme sans vérificationnisme ».

conception sémantiquement réaliste pour un certain domaine, mais qui adopte une ontologie antiréaliste pour les objets de ce domaine (par exemple en éthique) ; ou inversement une conception sémantiquement antiréaliste mais néammoins réaliste ontologiquement parlant[1]. Et il n'y a pas de raison pour que ce qui vaut pour un domaine d'entités (morales, mathématiques, etc.) vaille pour un autre. C'est l'un des effets du minimalisme quant à la vérité et à la signification. Il faudrait encore préciser bien des choses pour que ce réalisme minimal soit plus que programmatique[2]. Il n'entre pas dans mon propos, ici, de préciser quelle serait la meilleure version de ce réalisme minimal. J'ai seulement voulu soutenir que si *une* version de cette position est correcte, elle permet de répondre de manière satisfaisante au défi antiréaliste.

1. C'est ce que propose Wright au sujet des énoncés mathématiques dans Wright, 1983.
2. D'autres conceptions de ce type ont été proposées par Johnston, 1991 et par Pettit, 1990.

6

Réalisme et holisme

> What we call the beginning is often the end
> And to make an end is often to make a beginning
> The end is where we start from. And every phrase
> And sentence that is right (where every word is at home,
> Taking its place to support the others...)
> Every phrase and every sentence is an end and a beginning.
>
> T. S. Eliot, *Little Gidding*.

6.1. Ni correspondance ni référence

En montrant, dans le chapitre précédent, que le réalisme n'était pas une doctrine unifiée, et en définissant une position, le réalisme minimal, qui satisfait à un certain nombre des critères de l'antiréalisme, nous avons, dans une large mesure, désarmorcé la critique antiréaliste de la conception vériconditionnelle de la signification. Mais cela ne nous dit pas si le réalisme davidsonien répond de façon satisfaisante aux critères que nous avons mis en avant en définissant le réalisme minimal. Je soutiendrai ici que la position de Davidson est *une sorte* de réalisme minimal, même si la nature de ce réalisme, et les arguments sur lesquels il repose, diffèrent de façon importante des positions que nous avons examinées jusqu'à présent. Comme on peut s'y attendre, les différences tiennent principalement au poids qu'il accorde à l'interprétation radicale et à la nature spécifique de son holisme.

Pour voir comment ces problèmes se posent, le plus simple est de revenir à nouveau à la conception tarskienne de la vérité[1]. Davidson admet, avec

1. Je suivrai ici de plus ou moins près ce qui constitue à ce jour la dernière présentation par Davidson de ces questions, dans ses *Dewey Lectures* (Davidson, 1990).

la plupart des interprètes de Tarski[1], que ce dernier n'a pas *défini* le concept de vérité, au sens philosophique de cette notion : bien qu'il ait montré comment définir un *prédicat* de vérité pour chaque langage formel parmi un ensemble de langages de ce type, il n'a pas expliqué ce que ces langages avaient en commun. Il n'a pas plus expliqué ce concept, comme on le dit parfois, en termes de correspondance (nous allons voir dans un instant pourquoi)[2]. Mais il a montré, avec son emploi de la Convention T, que le prédicat « vrai » devait satisfaire à certains critères formels, en particulier celui de décitation. En d'autres termes, Tarski semble avoir justifié par avance ce que nous avons appelé le déflationnisme. Mais il y a, comme on l'a vu, deux façons d'interpréter ce fait, que Davidson résume très bien : ou on l'interprète comme « montrant que [Tarski] n'a pas représenté certains aspects essentiels du concept de vérité », ou on l'interprète comme « montrant que le concept de vérité n'est pas aussi profond et intéressant que beaucoup de gens l'ont pensé » (1990 : 288). La seconde interprétation est celle des auteurs qui entendent s'en tenir aux platitudes déflationnistes (le *quiétisme*). La première est celle de ceux qui acceptent ces platitudes, mais pour qui elles n'épuisent pas tout ce qu'il y a à dire sur la vérité. Davidson est clair sur le fait qu'il rejette la seconde position, et adopte la première (*ibidem*). La question est donc de savoir ce qu'il y a pour lui de *plus* à dire sur la vérité, y compris et surtout quand on adopte comme lui la thèse selon laquelle une théorie de la vérité de type tarskien doit pouvoir faire office de théorie de la *signification*. Il y a ici, comme on l'a vu, deux sortes de développements possibles, les uns dans le sens d'un réalisme externe, les autres dans le sens d'un vérificationnisme. Il importe de voir pourquoi Davidson les rejette tous deux.

Tout d'abord ce « quelque chose de plus » ne peut pas consister en une analyse de la vérité comme *correspondance*. Il y a deux manières de formuler cette idée. La première consiste à définir la vérité comme correspondance avec un certain type d'entités dans le monde, des faits ou

1. Dummett, 1959 ; Field, 1972, 1987 ; Soames, 1985 ; Etchemendy, 1988 ; Horwich, 1990.

2. Je fais ici allusion à, par exemple, l'interprétation de Popper, qui considère que la conception sémantique de la vérité de Tarski « justifie » le concept de correspondance. Cf. Engel, 1989, chap. v.

des états de choses. Mais Davidson (1969) soutient que c'est impossible. Supposons en effet que nous définissions la vérité au sens de Tarski, sur la base de la notion de *satisfaction* : une phrase du langage-objet est vraie si et seulement si elle est satisfaite par toute *suite* d'objets sur lesquels portent les variables de quantification du langage-objet. On peut dire en ce sens que les phrases vraies « correspondent » à la réalité, dans la mesure où le prédicat de satisfaction établit bien une certaine correspondance entre les variables et les *objets* qui figurent dans les suites. Mais ces objets ou les suites d'objets que sont les suites dans une sémantique tarskienne ne sont pas pour autant des « faits » au sens où l'entend habituellement une théorie de ce genre d'entités, c'est-à-dire des complexes d'individus, de propriétés, et de relations désignés par les phrases. Par exemple une phrase comme « Dolorès aime Dagmar » sera satisfaite par les deux individus Dolorès et Dagmar (dans cet ordre, une suite étant ordonnée), mais ces deux individus ne composent pas à eux seuls un « fait » : pour qu'il y ait un fait, il faudrait que la relation de satisfaction s'étende, pour ainsi dire, au prédicat « aime » de la phrase qui les « unit ». Mais c'est précisément ce que nous n'avons *pas* dans une théorie tarskienne, puisque la notion primitive est celle d'un prédicat ou d'une phrase ouverte satisfaite par des suites en vertu d'une fonction qui relie ces prédicats ou phrases ouvertes à des entités dans les suites : les prédicats eux-mêmes ne reçoivent pas de « correspondants » dans le monde, contrairement à ce que suppose une théorie correspondantiste des « faits ». Encore moins peut-on parler de *complexes* d'objets et de propriétés, unis d'une manière ou d'une autre, qui pourraient correspondre aux phrases (1969 : 48, *84*). A ceci s'ajoute une seconde difficulté, sur laquelle Davidson s'étend à plusieurs reprises. Selon un argument célèbre, dû à Frege et à Church, et appelé souvent « argument du lance-pierre », si deux phrases ont la même valeur de vérité, elles doivent avoir la même référence[1]. Une application directe de cet argument au cas qui nous occupe conduit à soutenir que si deux phrases ont la même valeur de vérité elles doivent désigner le même fait, et par conséquent que toutes les phrases équivalentes en valeur de

1. Cf. en particulier Davidson, 1967, 1969, 1967a, 1967c.
Cet argument est connu également sous le nom d'« argument du lance-pierre ». Je l'ai analysé dans Engel, 1989 : 18-20, et chap. 5.

vérité désignent le même fait, ou le Grand Fait. Ainsi supposons que l'on dise que l'énoncé que Naples est bien plus au nord que Red Bluff est vrai parce que cet énoncé correspond à un fait. On peut présumer que ce fait est le fait que Naples est bien plus au nord que Red Bluff. Mais ce peut être aussi le fait que Red Bluff est bien plus au sud que Naples, ou encore le fait que Red Bluff soit bien plus au sud que la plus grande ville italienne située à cent kilomètres d'Ischia, ou même le fait que la plus grande ville italienne située à cent kilomètres d'Ischia et telle que Londres est en Angleterre (par substitution *salva veritate* des expressions coréférentielles), car toutes les phrases auxquelles sont supposés correspondre ces faits sont extensionnellement équivalentes (1969 : 41, 75). Nous n'avons ainsi plus aucun moyen de distinguer l'idée selon laquelle une phrase est vraie si elle correspond à *un* fait de l'idée selon laquelle cette phrase correspond à *tous* les faits. Toutes les phrases vraies correspondent aux mêmes entités. Davidson tient cela comme une réduction à l'absurde de la théorie de la vérité comme correspondance à des faits. Cependant cela ne montre pas que *toute* théorie de ce type soit vouée à l'échec. On peut rejeter l'argument de Frege-Church de diverses manières, et chercher à construire une théorie récursive de la notion de fait qui permette d'individualiser de façon plus fine des entités de ce type. Mais il est clair que ces types de conceptions ne reposeront plus sur les concepts d'une théorie extensionnelle de la vérité tarskienne, et par conséquent que Davidson ne peut pas s'appuyer sur elles s'il veut maintenir son projet[1]. Mais même si l'on acceptait ce genre de révisions, il existerait des raisons indépendantes des difficultés propres à la notion de fait de rejeter une théorie de la vérité comme correspondance.

Ce sont ces raisons que Davidson avance, en second lieu, quand il examine la proposition de Field (1972) de modifier la théorie de la vérité tarskienne en l'augmentant d'une *théorie causale de la référence*. Field remarque qu'une définition de la vérité de type tarskien procède de façon purement *stipulative* en énumérant, sous forme de listes de clauses, pour un langage *L* particulier, les références des prédicats ou des noms du langage

1. Voir, par exemple Barwise et Perry, 1983 ; Taylor, 1985 ; Forbes, 1987, 1990. Taylor, 1985 montre qu'un ajout seulement minimal de notions intensionnelles à la construction tarskienne est nécessaire pour conduire à la conception des « états de choses » qu'il envisage.

(comme le fait la théorie T3 du § 1.3.4). Ces spécifications peuvent être considérées comme des *définitions partielles* des prédicats « vrai » et « dénote », qui éliminent ces prédicats. Ainsi, dans un langage *L* contenant seulement quatre termes, *a*, *b*, *c*, et *d*, on pourrait donner une définition de la référence de la forme :

> *x* dénote *y* dans *L* ssi *x* est *a* et *y* est Marx, ou *x* est *b* et *y* est Engels, ou *x* est *c* et *y* est Lénine, ou *x* est *d* et *y* est Mao (où « *x* » est une variable métalinguistique sur des noms de *L*, et où *y* une variable sur des individus)

et pareillement on pourrait donner des définitions de « satisfait » (pour les prédicats) et de « vrai » (pour les phrases). On aurait défini ces notions *pour L*, mais on n'aurait pas pour autant montré comment on peut définir *en général* ces notions pour *tout L*. Autrement dit, Tarski a bien montré comment on pouvait réduire le concept sémantique de vérité au concept sémantique de satisfaction (ou, dans la terminologie de Field, de « dénotation primitive »), mais il n'a pas produit ce qui, selon Field, devrait constituer une authentique définition d'une notion sémantique, c'est-à-dire une définition qui ne serait plus elle-même en termes *sémantiques*[1]. Putnam ajoute que des définitions partielles de ce type, étant purement stipulatives, réduisent les phrases-T correspondantes à de pures *tautologies*, qui deviennent des conséquences analytiques des stipulations initiales, et peuvent ainsi difficilement figurer au titre de vérités empiriques testables[2]. Cet argument est implicitement dirigé contre Davidson, puisque celui-ci entend précisément leur faire jouer ce rôle. Mais l'argument manque sa cible, parce que Davidson n'entend justement pas limiter ce qu'il appelle une théorie empirique de la vérité à l'emploi de définitions partielles de ce genre.

Que serait, d'après Field, une véritable définition des concepts sémantiques de vérité et de référence ? Ce serait une définition qui, à la différence des définitions partielles de Tarski, permet d'appliquer le concept de vérité à des cas nouveaux, et qui ne nous donne pas seulement l'*extension* du prédicat de vérité. Pour cela, il faut proposer une réduction du concept

1. Field, 1972. Cf. également Blackburn, 1984 : 265-273 ; Engel, 1989 : 131-133.

2. Putnam, 1983a, cf. aussi pour des versions du même reproche, Soames, 1984 ; Etchemendy, 1988. Cf. Davidson, 1990 : 288-293.

même de référence ou de dénotation en termes *physicalistes* ou naturalistes, c'est-à-dire ne faisant pas appel à des notions sémantiques ou intentionnelles, comparable à la définition de concepts comme celui de valence en chimie[1]. Field ne donne pas beaucoup de précisions sur ce à quoi pourrait ressembler une telle réduction, et se contente de suggérer qu'une théorie causale de la référence comme celle proposée par divers auteurs dans le cas des noms propres et des termes d'espèce naturelle serait un premier pas[2]. Néanmoins la théorie causale de la référence, sous sa version kripkéenne ou putnamienne, peut difficilement passer pour une réduction d'un concept sémantique à un concept non sémantique, car elle utilise la notion d'*intention* de référence. Des théories causales de la référence plus prometteuses en ce sens seraient celles de Stampe (1977) ou de Dretske (1981), qui font appel à des notions comme celles de contenu *informationnel* de représentations mentales ou sémantiques, ou celle de Fodor (1987, 1990), qui emploie l'idée d'une *covariance* entre des symboles (mentaux ou linguistiques) et des états du monde, ou encore celles de Millikan (1984) et de Dretske (1988) qui recourent à l'idée d'une relation *téléofonctionnelle* entre les représentations et ce qu'elles représentent. Une théorie causale de la référence en ce sens impliquera une théorie causale de la vérité, et donc une théorie de la vérité comme correspondance, dans la mesure où les relations causales entre les représentations et les choses qu'elles représentent impliqueront des relations de correspondance, sous forme de *lois*. Une bonne partie des arguments adressés par ces théoriciens à une conception comme celle de Davidson dépend des formulations précises de ces diverses théories causales, et il n'entre pas dans mon propos de les analyser[3]. Mais quel que soit le type de théorie causale proposée, elle devra faire face à un agenda assez lourd, et montrer au moins trois choses. En premier lieu, elle devra montrer que, pour élucider la signification d'une représentation linguistique (ou mentale), seule la dimension de la *référence* des signes suffit, et qu'il n'est pas nécessaire de faire appel à leur *sens*, ce qui implique qu'elle puisse répondre à des problèmes traditionnels comme le problème frégéen des termes singuliers coréférentiels dans

1. Field, 1972. Cf. Putnam, 1978 ; Engel, 1989 : 132.
2. Cf. Kripke, 1972 ; Putnam, 1975. Cf. Engel, 1985.
3. J'ai examiné certaines de ces théories dans Engel, 1992, chap. 5.

des contextes intensionnels ou opaques sans recourir à la notion de sens (fregéenne ou non). En second lieu, elle devra montrer que la définition causale proposée est *non circulaire*, c'est-à-dire que les notions supposées définir les notions sémantiques ou intensionnelles de signification et de référence, comme celles de « covariance » ou de « fonction propre », ne *présupposent* pas elles-mêmes ces notions ou des notions voisines. Enfin, elle devra être *unique*, au sens où elle devra montrer que les relations (ou « chaînes ») causales systématiques ou nomologiques qui peuvent, selon ce type de théorie, être établies entre les signes et les choses du monde qu'ils « représentent » sont d'une part *effectivement* nomologiques et d'autre part uniques, au sens où un seul type de relation causale doit correspondre à un seul type de signification. Une théorie causale doit, par conséquent, soutenir que les liens causaux sont non seulement nécessaires, mais suffisants, pour assurer le caractère *déterminé* des représentations et des significations[1]. Notons que ces réquisits ressemblent à ceux que Davidson lui-même formule pour une théorie adéquate de la signification (§ 1.1). Mais alors qu'une théorie causale de la référence ou de la signification présuppose que tous ces réquisits peuvent être satisfaits, Davidson soutient au contraire qu'ils ne peuvent pas l'être. Tout d'abord, il ne prétend pas réduire une théorie de la signification à une théorie de la référence. Comme on l'a vu au sujet des critiques de Dummett, une théorie de la signification n'est pas « modeste » en ce sens. Même s'il ne fait pas appel directement à la notion frégéenne de *sens*, ou de *Sinn*, Davidson est conduit à distinguer la dimension de la référence d'un signe de la dimension de son sens. La différence entre sa conception et celle d'une théorie de type fregéen est qu'il refuse d'assigner des *Sinne* ou des « modes de présentation » *individuels* aux noms propres et aux autres expressions. Pour « faire office » (au sens envisagé au § 1.5) de théorie du sens, une TS doit être holistique, et admettre qu'il n'est pas possible d'assigner *un* sens ou mode de représentation unique et fixe à chaque signe, mais que ce sens dépendra des *autres* assignations que fera l'interprète à un *ensemble* de signes et de

1. On trouve un bon exposé de cet agenda dans les conditions que Fodor lui-même impose à sa propre théorie du contenu, dans Fodor, 1990. Ramberg, 1989, chapitre 3, donne également un bon exposé de ces conditions, et des raisons pour lesquelles elles ne peuvent être remplies selon Davidson.

phrases où ils figurent. Ensuite Davidson considère comme impossible l'idéal de non-circularité (même s'il l'admet comme régulateur pour une théorie sémantique) : il soutient qu'on ne peut éliminer totalement d'une théorie (générale) de la signification des notions comme celle de signification, ou des notions intentionnelles comme celles de croyance, de désir, ou de préférence. Enfin Davidson considère qu'une théorie de l'interprétation maintiendra toujours un certain degré d'indétermination. L'idéal même de relations *nomologiques strictes* entre signes et objets du monde est illusoire (même si, comme on l'a vu, il est possible de parler de lois *ceteris paribus*, ou des phrases-T d'un interprète comme de *lois*). Chacune de ces « réponses » à la proposition d'une théorie causale repose évidemment sur une pétition de principe en faveur de la conception davidsonienne, et en ce sens elles ne montrent en rien que cette proposition soit vouée à l'échec. Tant qu'on n'a pas montré de façon définitive qu'une théorie de type causal est possible ou impossible, il est probable que les débats extrêmement virulents qui ont agité la philosophie du langage et de l'esprit sur ces sujets continueront. Mon propos n'est pas ici d'établir l'un ou l'autre. Je voudrais seulement essayer de montrer, à la suite de Davidson, que les hypothèses sur lesquelles il s'appuie pour douter de la possibilité sont plausibles et qu'elles n'ont pas les conséquences dramatiques ou inacceptables qu'y voient ses critiques. Ces hypothèses reposent toutes, à un degré ou à un autre, sur le type de holisme inhérent à sa conception de la signification. Je me contenterai d'abord d'exposer les divers arguments où intervient ce holisme, sous diverses formes, dans la défense du « réalisme » proposé par Davidson, et je n'envisagerai la notion de holisme pour elle-même qu'à la fin de ce chapitre.

Pourquoi, selon Davidson, une théorie causale de la référence comme celle que propose Field est-elle impossible ? Avant tout pour la même raison que celle pour laquelle une théorie « substantielle » et moléculariste de la signification telle que la propose Dummett l'est. Comme nous le savons, ce n'est pas le molécularisme en lui-même que Davidson rejette. Au contraire une théorie-T *doit* être moléculaire, car elle doit montrer que le sens des expressions complexes d'un langage dépend du sens des expressions simples qui les composent. En ce sens il n'y a pas d'autre choix, pour une théorie sémantique, que de procéder, pour employer

l'expression de Blackburn (1985 : 273), « de bas en haut » (*bottom up*), en décomposant les phrases en prédicats, en noms et en autres catégories grammaticales, et en montrant comment des objets sont dénotés par des noms et satisfont les prédicats, rendant ainsi vraies les phrases dans lesquelles ils figurent. Une théorie de la vérité qui ne révélerait pas une telle structure compositionnelle ne serait pas une théorie de la vérité. Mais une chose est la *structure* de la théorie elle-même, et ce que cette théorie explique au moyen de notions comme celle de prédicat, de nom ou de satisfaction, et autre chose est la *confirmation* ou l'application de cette théorie. Ce que le théoricien « causaliste » de la référence suppose, quand il propose de rattacher des unités linguistiques à des morceaux de la réalité ou à des morceaux de comportement, c'est que nous ayons d'une part une caractérisation suffisamment précise de ces unités (par la syntaxe) et d'autre part une caractérisation indépendante des conditions dans lesquelles ces unités peuvent être dites signifier ou faire référence à des portions de réalité non linguistique. Mais nous n'avons ni l'un ni l'autre. Davidson ne nie pas que nous ayons besoin d'une caractérisation syntaxique des expressions. Mais il soutient que nous ne pouvons pas donner de fonction sémantique à celles-ci indépendamment du rôle qu'elles jouent dans des *phrases*. C'est ici qu'intervient le principe frégéen de contextualité, ou ce que nous avons appelé le *holisme de la phrase*. Or une phrase n'est interprétable que relativement à la croyance d'un locuteur qu'elle est vraie (son tenir-pour-vraie cette phrase) et relativement aux intentions du locuteur de l'utiliser en vue d'énoncer quelque chose de vrai ou de faux, c'est-à-dire de faire ou non une assertion. Par conséquent on ne peut pas abstraire la phrase de cette caractérisation de son usage par un contexte d'attitudes propositionnelles et d'objectifs, pour essayer de voir ce qu'elle signifie, ou ce à quoi elle fait référence, indépendamment de ces attributions. Et ces attributions d'attitudes ne peuvent pas elles-mêmes se faire indépendamment de l'attribution d'*autres* attitudes. Deux principes jouent ici. Le premier est que le point où le langage peut rencontrer la réalité ne peut pas être situé au niveau des expressions simples ou des mots qui composent les phrases, mais seulement au niveau des phrases elles-mêmes. Cela implique que la première valeur sémantique que doit tester une théorie de la signification est la vérité, et non pas la référence. Une théorie de la réfé-

rence ne peut avoir qu'un rôle *dérivé* par rapport à une théorie de la vérité pour un langage. Le second principe est celui de l'interdépendance des croyances et des significations (§ 2.1) : on ne peut dire ce qu'une phrase signifie (et en l'occurrence comment ses termes ont une référence) indépendamment des attitudes de ceux qui l'emploient. C'est pourquoi, selon Davidson, on ne peut espérer expliquer la référence « directement en termes non linguistiques »[1]. La méthode atomiste et moléculariste, qui consiste à expliquer progressivement le sens d'une phrase à partir de la référence de ses parties composantes, « de bas en haut », ou, selon l'image de Davidson lui-même, par « empilage de blocs », successifs est donc vouée à l'échec. Mais inversement la méthode que Davidson propose, qu'il appelle « méthode holistique », et qui consiste à aller de « haut en bas » (*top down*, dans la terminologie de Blackburn) peut-elle nous fournir le degré d'analyse et de décomposition suffisants des phrases en constituants sémantiques ? Si le sens des phrases dépend nécessairement d'une « structure d'ensemble », et les constituants de leur « récurrence » à l'intérieur de cette structure, comment pourrons-nous jamais isoler ces constituants et parvenir à la structure moléculaire désirée ? C'est, comme on l'a vu, l'un des reproches que Dummett adresse au holisme de Davidson. Et ce dernier admet ce point, puisqu'il parle d'un « dilemme » entre d'un côté la théorie des blocs empilés qui lui paraît impossible et de l'autre la méthode holiste qui risque d'interdire toute prise sur la structure du langage (1977 : 221, *318*). La solution du dilemme consiste à abandonner la première branche de l'alternative et à choisir la seconde, mais en niant que la méthode holistique ait la conséquence désastreuse alléguée :

> J'ai admis qu'une théorie de la vérité à la Tarski n'analyse ni n'explique le concept pré-analytique de vérité pas plus que le concept pré-analytique de référence : au mieux, elle donne l'extension du concept de vérité pour tel ou tel langage doté d'un vocabulaire primitif fixe. Mais cela ne montre pas qu'une théorie de la vérité absolue ne peut expliquer la vérité de phrases individuelles sur la base de leur structure sémantique ; tout ce que cela montre, c'est que les caractéristiques sémantiques des mots ne peuvent pas être tenues pour fondamentales dans l'interprétation de la théorie. Ce qu'il faut pour résoudre le dilemme de la référence, c'est une distinction entre l'explication *dans* la théorie et l'explication *de* la théorie. Dans la théorie, les conditions de vérité d'une phrase sont spécifiées en postulant une structure et des concepts sémantiques comme ceux de satisfaction ou de référence. Mais lorsqu'il s'agit d'interpréter la théorie dans son ensemble, c'est la notion de vérité, appliquée à des phrases closes, qu'il faut relier à des fins et à des activités humaines. L'analogie avec la physique est évidente : nous expliquons les phénomènes macroscopiques en postulant une structure fine inobservée. Mais la théorie est testée au niveau macroscopique. Parfois, évidemment, nous avons assez de chance pour trouver d'autres preuves, plus directes, de la structure que nous avons postulée à l'origine ; mais cela n'est pas essentiel à l'entreprise. Je suggère que les mots, les significations des mots, la référence et la satisfaction sont des principes que nous avons besoin de poser pour réaliser une théorie de la vérité (1977 : 221-222, *320-321* ; cf. aussi 1979 : 236, *340*).

1. Davidson, 1977 : 220-221, *319-321*.

Cette distinction entre l'explication *interne* à la théorie, et l'explication *de* la théorie recoupe celle qui nous est à présent familière entre une théorie-T et la théorie de l'interprétation radicale, ou entre ce que Dummett appelle les « deux composantes » de la conception davidsonienne. La comparaison entre une théorie de la vérité et une théorie physique nous est aussi familière, sous la forme de l'idée qu'une théorie-T joue, comme la théorie bayésienne de la préférence dans l'interprétation « généralisée » du langage et de l'action, le même rôle que la théorie de la mesure en physique (§ 2.4).

Une théorie causale de la référence ne pourrait donc être au mieux que circulaire. Mais Davidson soutient aussi qu'une théorie de ce genre ne serait pas unique. Ici son argument (1979) dépend étroitement de la thèse quinienne de l'indétermination de la traduction, ou plus exactement d'une version de cette thèse, l'*inscrutabilité de la référence*. Supposons qu'il y ait des relations causales entre les mots, ou leurs usages, et des objets. Dans ce cas, une phrase comme « Meaulnes est grand » peut être vraie si, entre autres, une énonciation du mot « Meaulnes » dans ce contexte verbal est reliée causalement avec l'individu Meaulnes. Mais avons-nous par là déterminé que le schème de référence choisi n'est pas arbitraire ? Davidson utilise ici un argument qui a été fourni par Field lui-même[1]. Supposons qu'il y ait ce que l'on peut appeler une permutation de l'univers, c'est-à-dire une application bi-univoque de tout objet sur un autre.

1. Field, 1974 ; cf. aussi Wallace, 1977.

Supposons que Φ soit une telle permutation. Si nous avons un schème satisfaisant pour un langage parlant de cet univers, nous pouvons produire un autre schème de référence utilisant cette permutation : chaque fois que, dans le premier schème un nom désigne un objet x, dans le second schème il désigne $\Phi(x)$; chaque fois qu'un prédicat dans le premier schème désigne chaque x qui est F, dans le second schème, il désigne chaque x tel que $F\,\Phi(x)$. On voit que les conditions de vérité que le second schème assigne à une phrase seront dans chaque cas équivalentes aux conditions de vérité assignées à cette phrase par le premier schème. Mais les deux schèmes ne seront pas équivalents, comme le montre l'exemple suivant. Supposons que tout objet a une ombre et une seule. On peut prendre alors Φ comme exprimant le prédicat « est une ombre de ». On peut alors avoir un premier schème dans lequel le nom « Meaulnes » désigne Meaulnes, et le prédicat « est grand » désigne des choses grandes. Dans le second schème, nous tenons « Meaulnes » comme désignant l'ombre de Meaulnes, et « est grand » comme désignant les ombres des choses grandes. La première théorie nous dit que « Meaulnes est grand » est vrai ssi Meaulnes est grand ; la seconde nous dit que « Meaulnes est grand est vrai ssi l'ombre de Meaulnes est l'ombre d'une chose grande ». Les conditions de vérité sont équivalentes. La référence est inscrutable en ce sens. Voyons maintenant si elle peut l'être quand nous supposons que nous connaissons une relation causale appropriée entre mots et choses, et soit C (x, y) cette relation causale entre un mot x et un objet y. Une théorie causale nous dit que « Meaulnes » désigne Meaulnes seulement si C (« Meaulnes », Meaulnes), alors qu'une autre théorie nous dit que « Meaulnes » désigne Meaulnes seulement si C (« Meaulnes », Φ (Meaulnes)). Les deux théories sont distinctes, et les termes n'ont pas la même extension, mais elles sont équivalentes dans leurs conditions de référence. Il s'ensuit que :

> Aucune théorie causale, ni aucune autre analyse « physicaliste » de la référence, n'affectera nos arguments en faveur de l'inscrutabilité de la référence, pas du moins tant que nous admettrons qu'une théorie satisfaisante est une théorie qui produit une explication acceptable du comportement verbal et des dispositions verbales. Car l'on peut toujours saisir de manière équivalente les contraintes qui pèsent sur la référence et la causalité (ou tout ce qu'on voudra) par différentes

> manières de faire correspondre les mots avec les objets. L'interprète de l'inventeur de schèmes sera donc en mesure de dire que les schèmes de l'inventeur sont différents l'un de l'autre, mais il ne sera pas en mesure de désigner une seule et unique manière de faire correspondre les mots et les objets de l'inventeur. Il s'ensuit que l'inventeur ne peut avoir utilisé de mots ayant déterminé un schème unique. La référence reste inscrutable (1979 : 237, *341*).

En d'autres termes, une théorie moléculaire, causale ou non, sera toujours face à l'inscrutabilité de la référence des termes qu'elle tient comme les pierres de construction de la signification. Si ce fait est irréductible, il vaut mieux vivre avec lui, et en admettre les conséquences.

Dummett soutient que ces conséquences sont insupportables, parce que l'inscrutabilité entraîne qu'un locuteur ne pourra jamais connaître la référence de ses termes. Mais la conclusion ici est seulement qu'il y aura toujours une indétermination dans l'interprétation. Cette conclusion apparaît aussi inacceptable au partisan d'une théorie substantielle de la signification. Mais nous avons vu aussi que cette conclusion est loin d'être aussi radicale qu'il n'y paraît, parce que la procédure de Davidson réduit considérablement la portée de l'indétermination en question.

La réponse de Davidson à Field nous indique trois choses importantes. Tout d'abord elle nous indique que le holisme qu'adopte Davidson est bien *méthodologique* et non pas « constitutif » : le but d'une procédure d'interprétation est bien de parvenir à des assignations de signification qui révèlent une structure moléculaire, en postulant une théorie qui soit elle-même moléculaire, et en la testant indirectement à partir du comportement observable. Les principes de la procédure interprétative sont holistiques (en un sens qui reste encore à préciser), mais rien n'indique que les *produits* de la procédure, c'est-à-dire les assignations de signification elles-mêmes, le soient. Sans quoi on ne pourrait tout simplement pas parler d'interprétation. En second lieu, le fait que l'on ne puisse pas avoir une caractérisation de la référence indépendamment d'une description du comportement linguistique des agents, et une caractérisation de ce comportement indépendamment de leurs attitudes, montre que l'on ne peut pas adopter, selon Davidson, le type de point de vue « externe » que préconisait Dummett, car ce point de vue implique précisément que l'on ait des caractérisations indépendantes de ce type. En ce sens, Dummett

a raison de voir dans la conception davidsonienne une conception « modeste » de la signification, bien qu'il ait tort de supposer que cette conception réduit une théorie de la signification à une théorie de la référence. Enfin et en troisième lieu, nous obtenons aussi une réponse à la question que nous posions initialement dans cette section : qu'y a-t-il de *plus* dans une théorie de la vérité que les platitudes déflationnistes ? Le *plus* en question ne peut pas être une conception de la vérité comme correspondance, et en ce sens Davidson rejette implicitement le réalisme externe sur lequel se fonde ce genre de conception. Ce qu'il y a de plus dans une théorie de la vérité que les platitudes ou les stipulations, c'est une conception de la manière dont la notion de vérité se relie à des attitudes et à des activités humaines, c'est-à-dire une théorie empirique de la manière dont les locuteurs peuvent donner un sens et des conditions de vérité à des phrases qui ne sont en elles-mêmes que des suites de signes ou de sons[1]. Il y a ici une analogie étroite et intéressante — que nous avons déjà exploitée et à laquelle nous reviendrons encore — entre la théorie du langage et la théorie de l'action selon Davidson. Une théorie de l'action tient les actions comme des événements individuels, qui peuvent être décrits en termes physiques et qui *doivent* être décrits en termes mentaux ou intentionnels s'il faut les compter comme des actions. Les événements sont des entités réelles, existant indépendamment des descriptions que nous en donnons, et appartenant à l'ordre causal du monde, mais ils ne prennent leur sens comme *actions* que dans le contexte des descriptions psychologiques (en termes de raisons) que nous en donnons. De même les mots et les phrases d'un langage sont des symboles physiques, qui entrent dans des relations physiques avec d'autres entités physiques (des états du monde, des événements physiques dans le corps des agents), mais ils ne prennent sens que quand nous les décrivons dans le vocabulaire intensionnel de la signification, des attitudes et des buts humains. Un certain schème de description purement extensionnel (en termes d'événements, en termes de relation de dénotation entre noms et objets) est applicable dans les deux cas, de façon interne. Mais une autre description de ce schème est possible, qui l'explique en des termes intensionnels indépendants des concepts du premier.

1. Davidson, 1990 : 309-310.

6.2. *Arguments transcendantaux*

Si Davidson n'est pas un réaliste au sens du réalisme externe, ni au sens d'une théorie de la vérité-correspondance, en quel sens est-il un réaliste ? Il lui arrive d'endosser ce label explicitement :

> Si nous donnons une épistémologie correcte, nous pouvons être des réalistes dans tous les départements. Nous pouvons accepter des conditions de vérité objectives comme clef pour la signification, une conception réaliste de la vérité, et nous pouvons insister sur le fait que la connaissance est indépendante de notre pensée et de notre langage (1983 : 307).

Mais dans ses *Dewey Lectures* il admet que la thèse réaliste ne peut avoir de sens que si elle s'appuie sur une conception correspondantiste de la vérité, et qu'il a lui-même adopté cette étiquette pour indiquer qu'il rejetait toute forme de conception épistémique de la vérité, qu'il s'agisse d'une conception empiriste, d'une conception qui définisse la vérité en termes de cohérence entre les croyances, ou d'une conception vérificationniste, idéale ou non[1]. Selon Davidson, sa propre position n'est *ni* réaliste au sens où elle soutiendrait que la vérité est « radicalement non épistémique » *ni* antiréaliste au sens où la vérité serait « épistémique », mais elle ne vise pas à *concilier* le réalisme et l'antiréalisme, les deux positions étant pour lui « inintelligibles » (1990 : 298).

Ses arguments en faveur de cette thèse prennent tous leur point de départ dans la situation de l'interprète radical. La procédure d'interprétation radicale est supposée être la seule position à partir de laquelle on puisse avoir accès non seulement aux contenus de signification et de pensée des locuteurs d'un langage, mais également des contenus de toute pensée portant sur une quelconque réalité objective. L'interprétation radicale est la seule situation à partir de laquelle on peut envisager les conditions de possibilité d'une expérience possible, et elle est, en ce sens quasikantien, *transcendantale*. C'est ici que nous retrouvons ce que Davidson lui-même appelle l'« argument transcendantal » (cf. § 2.3) qui nous fait passer de

1. 1990 : 304-308.

la *présomption* de la vérité et de rationalité des contenus interprétées à la vérité et à la rationalité *effectives* de ces contenus, et qui fait de l'interprétation radicale une « épistémologie au miroir de la signification »[1]. Cet argument prend trois formes, la première celle d'une critique du relativisme et de la notion de schème conceptuel, la seconde celle d'une réfutation *a priori* du scepticisme, la troisième celle d'une détermination de l'ontologie à partir du langage.

(A) Dans son célèbre essai « Sur l'idée même de schème conceptuel » (1974), Davidson entreprend de ruiner le relativisme, entendu soit au sens où divers systèmes conceptuels seraient incompatibles ou incommensurables, soit au sens où divers systèmes de signification ou langages le seraient. Son argument n'est pas très clair, mais on peut considérer qu'il a quatre prémisses majeures. (a) Davidson commence par montrer que toute forme de relativisme doit reposer sur la distinction entre un *schème* ou un système de concepts ou de significations supposé *organiser* un certain donné, une certaine expérience ou un certain *contenu* d'une part, et ce donné, expérience, ou contenu d'autre part, supposé indépendant du schème qui l'organise. (b) Or il n'y a rien de plus, dans l'idée de schème conceptuel, que l'idée qu'un certain ensemble de phrases sont *vraies*, et il n'y a rien de plus, dans l'idée d'un système de significations ou d'un langage, que l'idée qu'un certain ensemble de phrases sont *tenues pour vraies* par des locuteurs, et par conséquent il n'y a rien de plus dans l'idée que deux schèmes seraient distincts que l'idée que des locuteurs tiennent certaines phrases comme vraies, alors que d'autres locuteurs tiennent un autre ensemble de phrases comme vraies, ou le même ensemble de phrases comme fausses. Ici l'un des prémisses de l'argument est (b') le rejet, à la suite de Quine, de la distinction entre énoncés analytiques et énoncés synthétiques, c'est-à-dire de la distinction entre un langage (un système de significations) et une théorie, c'est-à-dire un ensemble de phrases tenues pour vraies, et par conséquent de la distinction entre « vrai en vertu de la signification » et « vrai en vertu de l'expérience ». (c) Mais la distinction schème/

1. Davidson fait allusion à cet argument « transcendantal » dans le texte (1975 : 169, *250* déjà cité ci-dessus (§ 2.3) mais aussi dans 1973a : 72, *117* et 1974 et 1975.

contenu est incohérente, pour la même raison que celle pour laquelle une théorie de la vérité comme correspondance est incohérente : elle suppose que nous ayons d'un côté un certain nombre d'entités (des concepts, des significations) qui s'adapteraient à des entités particulières (des objets, des faits, des expériences sensorielles) ou à une totalité d'entités de ce type (le monde, l'expérience). Or

> Le problème est que la notion d'un accord avec la totalité de l'expérience, tout comme la notion d'un accord avec les faits, n'ajoute rien d'intelligible au simple concept d'être vrai. Parler de l'expérience sensorielle plutôt que des données, ou simplement des faits, exprime une conception de la source des données, mais elle n'ajoute pas une nouvelle entité à l'univers par rapport à laquelle on puisse tester des schèmes conceptuels. La totalité de l'expérience sensorielle est ce dont nous avons besoin pour autant que ce soient les seules données dont nous disposions ; et les seules données dont nous disposions sont simplement ce qui rend nos phrases vraies. Rien, cependant, aucune chose, ne rend les phrases et les théories vraies : aucune expérience, aucune irration de nos surfaces sensorielles, ni même le monde, ne peut rendre les phrases vraies. Que l'expérience prenne un certain cours, que notre peau soit chauffée ou blessée, que l'univers est fini, voilà des faits, si nous voulons employer cette expression, qui rendent vraies nos phrases et nos théories. Mais on peut en dire autant sans mentionner les faits. La phrase « Ma peau est chaude » est vraie si et seulement si ma peau est chaude. Ici il n'y a pas de référence à un fait, un monde, une expérience ou des données (1974 : 193-194, *283*).

(d) Il s'ensuit que quand on dit que deux schèmes sont incompatibles, ou que deux langages sont intraduisibles l'un dans l'autre, on soutient qu'il y a des ensembles de phrases qui sont *largement vrais* (pour des locuteurs à une époque donnée, ou pour des locuteurs appartenant à des « cultures » différentes) mais néanmoins intraduisibles. Mais nous ne pouvons pas comprendre la notion de traduction, et par conséquent celle de signification, indépendamment de celle de vérité. Et — c'est ici le point fondamental — l'idée même d'une traduction ou d'une interprétation d'une langue distincte de la nôtre doit reposer, en vertu du principe de charité, sur l'idée que la plupart des croyances de ceux que nous interprétons doivent être vraies, et que l'interprète ou le traducteur doit partager un ensemble de croyances vraies avec l'interprété. Il s'ensuit que nous ne pouvons pas, en vertu de la nature même de l'interprétation,

juger que d'autres locuteurs de notre langage, ou des locuteurs d'un autre langage, ont des croyances *radicalement* différentes des nôtres. Par conséquent nos langages *doivent*, en ce sens, être traduisibles. Par conséquent nos schèmes conceptuels ne peuvent pas être à la fois largement vrais et intraduisibles. Et si c'est le cas, alors (e) non seulement l'idée même de schèmes conceptuels ou de langages intraduisibles est incohérente, mais elle est fausse, et le relativisme, conceptuel ou sémantique, doit être faux également.

La prémisse (b) (et (b')) de l'argument semble manifester l'acceptation de la doctrine quinienne de l'indétermination de la traduction (IT). Mais Davidson s'écarte au contraire très fortement de l'orthodoxie quinienne. Tout d'abord, il assimile la position de Quine à une position empiriste et correspondantiste, dans la mesure où elle fait dépendre la vérité de nos croyances de leur ajustement à une expérience (par l'intermédiaire des phrases observationnelles). En ce sens, Quine peut encore parler du « contenu empirique » de nos phrases, et il maintient la distinction schème/contenu (il a, suggère Davidson, rejeté correctement les deux dogmes de l'empirisme, mais conservé un troisième dogme, qui est cette distinction même). Ensuite la thèse IT de Quine a une allure relativiste : elle implique, selon l'une de ses versions, que deux théories empiriquement équivalentes peuvent être vraies ou fausses en même temps, mais elle semble impliquer, selon une autre version, que la vérité et la référence sont relatives à une théorie d'arrière-plan, au sens où un locuteur qui accepte une théorie à un temps donné comme vraie, et une autre comme fausse, peut, à un temps ultérieur, accepter la première comme fausse et la seconde comme vraie[1]. Davidson accepte la première version (IT proprement dite, ou l'inscrutabilité de la référence au sens évoqué à la section précédente), mais il rejette la seconde, que Quine (1969) appelle la « relativité de l'ontologie ». Selon cette seconde version, la vérité, la référence, et l'ontologie sont relatives à une théorie d'arrière-plan ou à un langage, et deviennent en ce sens « immanentes » à ce langage ou cette théorie. Mais cette thèse, selon Davidson, ne suit pas de la première : si deux théories empiriquement équivalentes sont incompatibles, elles doivent être énoncées dans le *même* langage. Revenons aux deux schèmes équivalents

1. Davidson, 1979 : 232-234, 336-337 ; Davidson, 1990 : 306.

évoqués ci-dessus. L'un nous dit que « Meaulnes » désigne Meaulnes, l'autre que « Meaulnes » désigne l'ombre de Meaulnes. C'est ce fait qui nous induit à dire, comme Quine, que la référence n'a pas le même sens dans l'un et l'autre schème, et donc qu'elle doit être relativisée, ainsi que l'ontologie (une ontologie d'individus et une ontologie d'ombres d'individus). Mais ceci suggère qu'une fois que l'on a choisi (arbitrairement) un schème de référence, la référence a elle-même été fixée, relativement à ce choix. Mais elle ne l'a pas été et ne peut pas l'être, précisément parce qu'elle est inscrutable. C'est l'idée même que la référence et l'ontologie sont relatives à un langage ou à une théorie qui est problématique. Davidson admet bien que la vérité et la référence soient relatives à un langage, au sens trivial et inoffensif où les phrases sont vraies, et où les mots ont une référence, seulement relativement à un langage, et où un locuteur doit parler un langage ou un autre, mais il n'admet pas l'idée qu'elles soient relatives à un schème.

On dit souvent que Davidson n'a pas montré qu'il n'existe pas de schèmes conceptuels[1]. C'est correct. Dans la mesure où la notion de schème se réduit à celle d'un ensemble de phrases acceptées comme vraies, il est parfaitement possible, et banal, que de tels ensembles divergent, et qu'en ce sens il y ait des schèmes conceptuels. (Mais on ne voit plus très bien non plus à quoi peut servir l'idée de « schème ».) Ce que Davidson a montré est seulement que si l'on admet les principes de sa théorie de l'interprétation il ne peut pas y avoir une divergence *radicale* entre des schèmes, une « incommensurabilité » ou une intraduisibilité *totale*, au sens du relativisme conceptuel. Rien n'interdit des échecs partiels de la traduction.

Comment l'argument contre la notion de schème conceptuel se rattache-t-il au problème du réalisme ? En rejetant la notion relativiste de schème conceptuel et la dualité schème/contenu, Davidson rejette *à la fois* la thèse idéaliste selon laquelle la réalité serait relative à un schème *et* l'idée réaliste externe selon laquelle il existerait une réalité « non interprétée, quelque chose qui soit en dehors de tous les schèmes et de la science » (1974 : 198, *289*). Mais cela ne revient pas, selon lui, à abandonner la notion de vérité objective :

1. Cf. par exemple Blackburn, 1985, 60-61.

> Bien entendu la vérité reste relative au langage, mais c'est là quelque chose d'aussi objectif que possible. En abandonnant le dualisme du schème et du monde, nous n'abandonnons pas le monde, mais nous rétablissons un contact sans médiation avec les objets familiers qui, par leurs cabrioles, rendent nos phrases et opinions vraies ou fausses (*ibid.*).

En d'autres termes, selon Davidson, une forme de réalisme est possible selon laquelle nos phrases peuvent être vraies ou fausses au sujet d'une réalité objective, sans pour autant qu'il y ait une *confrontation* entre ces phrases et une réalité « non interprétée »[1].

(B) La seconde forme d'argument « transcendantal » qu'on trouve chez Davidson est sa critique du scepticisme radical dans « A Coherence Theory of Truth and Knowledge » (1983, cf. aussi 1982b). Le raisonnement est ici, au moins en apparence, très simple. Le scepticisme radical est entendu au sens usuel comme la thèse selon laquelle *toutes* nos croyances pourraient être fausses (par exemple si nous étions trompés par un Malin Génie, ou des cerveaux dans des cuves manipulés par un savant fou). Mais 1 / on ne peut pas interpréter une croyance sans présupposer d'une part (par le principe de charité comme cohérence PCR, cf. § 2.2) que cette croyance est reliée logiquement à un certain nombre d'autres croyances, et sans présupposer d'autre part (PCV) que la plupart des croyances interprétées sont vraies. Il s'ensuit 2 / que quiconque a des croyances et est capable d'interpréter des croyances doit supposer que la plupart des croyances sont vraies, et par conséquent 3 / que la plupart de nos croyances *sont* vraies et que le scepticisme radical est impossible. CQFD ! Mais notre réaction immédiate est que l'argument n'est pas valide, ou qu'il prouve trop. Je ne mettrai pas en question, pour le moment, le principe de charité. Ce qui paraît immédiatement incorrect dans l'argument est précisément la transition que nous avons déjà trouvée problématique, qui passe du fait que l'interprète *présume* que la plupart des croyances qu'il interprète sont vraies et rationnelles au fait que la plupart des croyances *sont* vraies. Cela n'implique-t-il pas une forme de vérificationnisme, selon laquelle ce qui apparaît être tel ou tel aux yeux d'un interprète doit

1. C'est ce qu'il résume dans Davidson, 1983, au moyen du slogan « correspondance sans confrontation », bien qu'il rejette, en 1990, la notion de correspondance.

être tel ou tel ? Davidson pourrait certes rejeter cette objection en disant qu'elle présuppose précisément ce qu'il conteste, à savoir l'idée (réaliste) qu'il pourrait y avoir un fossé irréductible entre ce que les choses sont et ce que nous pensons que les choses sont, fossé qui ne pourrait se justifier que par une conception réaliste « externe » de la vérité. Mais s'il rejetait cette présupposition, il devrait aussi rejeter son argument contre le scepticisme, et soutenir seulement qu'il a montré que la *présomption* du sceptique que toutes nos croyances pourraient être fausses est impossible. Ce serait une thèse bien plus faible que la « réfutation » annoncée, et on ne voit plus en quoi il y aurait un argument « transcendantal » ici[1]. S'il veut maintenir la version forte de la thèse, Davidson a autant intérêt à admettre l'intuition « réaliste » selon laquelle il peut y avoir une différence entre ce qu'un interprète tient pour vrai et ce qui est effectivement vrai, et faire face à l'objection qui lui impute une forme de vérificationnisme. L'objection, que Davidson formule lui-même (1983 : 317), est la suivante : ne pourrait-il y avoir une interprétation correcte d'un locuteur (interprète) par un autre en présence d'un ensemble de croyances *fausses* (et largement cohérentes) partagées par les deux ? La réponse de Davidson consiste à imaginer un interprète « omniscient », par conséquent différent d'un interprète « faillible » comme celui que nous envisageons. Etant ominiscient, cet interprète connaîtrait toutes les causes des croyances de l'interprète faillible et de celui qu'il interprète. Nous avons donc les croyances qu'un interprète omniscient nous attribuerait, et par conséquent nous devons, en tant qu'interprètes faillibles, être corrects et cohérents selon les critères objectifs de l'interprète omniscient. Il semble que l'invocation d'un interprète omniscient revienne à adopter précisément le point de vue « externe », ou divin, qu'on voulait précisément ne pas avoir à évoquer, ou revienne à l'affirmation de conditions d'acceptabilité idéales. La clef de l'argument repose sur l'idée de *causalité* qui intervient ici. C'est l'une des contraintes de l'interprétation radicale que l'interprète s'aide de ce qu'il tient comme les causes des croyances des agents qu'il interprète. La critique du scepticisme repose bien plus sur l'idée qu'un individu qui a des croyances qui sont généralement causées par son environnement exté-

1. Cf. Child, 1987 : 553.

rieur que sur la présomption de véridicité incorporée dans le principe de charité, ou plutôt ce qui justifie en dernier recours ce principe c'est précisément l'existence de cette interaction causale[1]. L'interprète peut se tromper sur les causes de ses croyances et de celles de l'interprété. Mais il ne peut pas se tromper totalement. On retrouve ainsi une forme beaucoup plus classique d'argument contre le scepticisme, et il n'est pas évident, tant qu'on n'a pas précisé la notion de causalité en jeu ici, que le scepticisme soit réfuté. Je reviendrai sur cette condition causale. Mais on peut déjà voir qu'elle ôte une partie de son caractère mytérieux à l'argument (B) : on présuppose que l'interprète fait déjà partie d'un monde objectif, qui cause ses croyances, et on ne dérive pas la vérité d'une majeure partie de ces croyances d'une présomption de l'interprète. Cette condition causale nous indique aussi en quoi il peut y avoir une réalité indépendante de nos croyances. Davidson donne cependant une autre raison de considérer *a priori* les croyances comme vraies. Il semble suggérer que la seule *cohérence* des croyances serait une justification pour leur vérité (1983 : 307). Il est clair que si c'est là son argument, il est peu concluant, car il est ouvert précisément à la critique que Davidson adresse lui-même ailleurs à une théorie cohérentiste de la vérité : des ensembles cohérents de croyances peuvent être faux[2]. La condition de cohérence en question ne peut fonctionner que si elle est reliée à la condition causale qui vient d'être mentionnée, et pour que l'on puisse être assuré avoir garanti quelque chose comme une connaissance, il ne suffit pas de montrer que « la plupart » des croyances doivent être vraies, mais aussi qu'elles sont *justifiées*. Il est douteux, par conséquent, que l'argument (B) prouve la fausseté du scepticisme radical. Il est beaucoup plus modeste, et procède d'une hypothèse réaliste, plutôt qu'il ne l'établit : ce que nous croyons être vrai au sujet du monde est déterminé par le monde lui-même.

(C) La troisième forme d'argument procédant des conditions de l'interprétation radicale à une conclusion sur la nature de la réalité est celle qu'on

1. Cf. Davidson, 1983 : « Ce qui interdit un scepticisme global quant aux sens est, selon moi, le fait que nous devons, dans les cas les plus simples et méthodologiquement les plus basiques, prendre les objets d'une croyance comme des causes de cette croyance. » Cf. aussi 1991b : 199 ; 1991c : 159.

2. 1990 : 305. Fodor et Le Pore, 1992 : 156, notent qu'en ce sens PCR et PCV sont équivalents.

trouve dans « La méthode de la vérité en métaphysique » (1977). Davidson soutient qu'en partageant un langage « on doit partager une image du monde qui est, dans une large mesure, vraie » (1977 : 199, *290*). Cette thèse n'est pas une répétition des arguments précédents, bien qu'elle repose aussi sur le principe de charité. La thèse porte plutôt sur ce qu'une analyse de la *structure* logique des phrases peut nous apprendre sur l'ontologie propre au langage, et sur ce que l'on peut en déduire quant à la structure de la réalité. Un interprète postule une telle structure comme une forme d'échelle ou de mesure. Si le système de mesure est, à un large degré, arbitraire (tout comme la température peut être mesurée en degrés Farhenheit ou centigrades) et indéterminé (les mesures ne sont pas uniques), il y a un invariant (§ 2.6). Cet invariant, dans le cas de l'analyse de la structure des phrases, est la forme logique. Quel que soit le degré d'arbitraire qui existe dans l'emploi d'une telle notion en sémantique, certains traits de base doivent être récurrents. Par exemple, l'analyse d'un grand nombre phrases d'actions, de phrases énonçant des relations causales et comportant une modification adverbiale révèle qu'elles impliquent une quantification implicite sur des entités particulières, des *événements*[1]. Nous avons en ce sens une raison de compter les événements comme des entités ultimes de l'« ameublement du monde », et de considérer qu'ils font partie des conditions de vérité des phrases en question. Davidson ne soutient pas qu'il suffise, pour établir l'existence de certaines entités, de se livrer à l'analyse du langage[2]. Mais il soutient que si une certaine ontologie permet de rendre vraies et d'interpréter une majorité de phrases de notre langage, alors nous pourrons dire que cette ontologie est effectivement la nôtre. Sans la réalité des événements, un bon nombre de ses doctrines sur la causalité, l'action et les rapports entre le physique et le mental seraient intenables. Ils font donc bien partie d'une réalité indépendante.

Nous pouvons voir mieux à présent quelle sorte de réalisme Davidson soutient. Il rejette le réalisme externe ou transcendant, qui présuppose une conception correspondantiste et une dualité du schème et du

1. Davidson, 1967, 1969 et 1980, *passim*. La thèse de Davidson est fortement contestée sur ce point. Cf. ci-dessus, § 1.4.

2. Cf. par exemple 1980 : xiv ; 1977 : 214, *311*.

contenu. Il rejette également une conception vérificationniste qui réduirait la vérité à l'assertabilité, et qui elle-même reposerait sur la dualité schème/contenu, puisqu'elle supposerait que nos énoncés puissent être justifiés par des expériences particulières. Le réalisme externe comme le vérificationnisme supposent que la vérité puisse être une « confrontation » (1983 : 307) avec des portions de réalité ou d'expérience particulières, phrase après phrase. Mais aucune confrontation directe de ce genre n'est possible. Nos phrases ne sont rendues vraies ni par des entités particulières, ni par une totalité d'entités ou d'expériences. Elles ne « rencontrent » même pas, comme le dit Quine, « le tribunal » de l'expérience en bloc. Les arguments contre les schèmes conceptuels et le scepticisme montrent que Davidson entend rejeter l'association du réalisme à l'idée de la fausseté possible de toutes nos croyances. Mais si la méthodologie de l'interprétation entraîne qu'il ne peut pas y avoir une distance inconcevable entre nos croyances et la réalité, et si l'on doit inférer que les choses sont telles ou telles parce qu'un interprète les tient comme telles ou telles, quel contenu peut bien avoir le réalisme de Davidson ? J'ai essayé de montrer qu'il n'y avait, dans les arguments « transcendantaux » de Davidson, aucune forme illégitime de vérificationnisme. Ce qui justifie, en dernier ressort, le principe de charité, c'est l'existence d'une relation causale générale entre nos croyances et la réalité. Quel que soit le degré de correction *a priori* de nos croyances, la réalité qui les cause demeure indépendante de celles-ci. Pour reprendre la distinction de Wright, Davidson assure l'objectivité de la vérité. Le problème demeure alors de savoir comment il peut assurer l'objectivité du jugement, c'est-à-dire en quoi nos énoncés peuvent porter sur un monde *objectif*.

6.3. Triangulation, communication et externalisme

La réponse à cette question est à nouveau à trouver dans les conditions de l'interprétation radicale. Une interprétation correcte investit nécessairement la personne interprétée d'un minimum de rationalité et de véracité. Mais le critère de cette rationalité et de cette véracité n'est pas purement relatif à l'interprète. Si ce dernier comprend celui qu'il interprète, ce dernier doit aussi partager ce même critère. Il est donc *interpersonnel* ou intersubjectif (bien qu'il ne soit pas *impersonnel* et transcendant : différents degrés de rationalité et de véracité sont possibles). Mais on peut demander : comment un critère intersubjectif peut-il être un critère objectif (1991c : 159) ? Pourquoi ce sur quoi les individus s'accordent devrait-il être vrai ? Nous retrouvons ici l'un(e) des critères ou des « marques » de la vérité de Wiggins et de Wright : la convergence. Comment est-elle possible et en quoi garantit-elle l'objectivité ? La réponse de Davidson, inspirée de celle de Quine, est que les humains classent des objets et des propriétés du monde en groupant des stimuli qu'ils reçoivent du monde extérieur comme similaires, et traitent ces stimuli comme similaires parce qu'ils produisent sur eux-mêmes et leurs semblables des réponses similaires. Ces structures de réponses peuvent s'expliquer par l'évolution et l'apprentissage. En ce sens, nous avons bien besoin d'une forme de théorie naturaliste des représentations, qui regroupe des structures comportementales et établit des régularités nomologiques. Mais le critère de la similarité des réponses ne peut pas lui-même être dérivé des réponses elles-mêmes, sans quoi il serait circulaire. Le critère de la similarité « ne peut venir que des réponses d'un observateur aux réponses de la créature » qu'il observe (1991c : 159). Cet observateur doit corréler les réponses d'un autre observateur aux objets et événements de son propre univers. Dans le cas de la communication verbale, une régularité du comportement verbal doit être corrélée aux objets que l'observateur perçoit dans son environnement. Il y a là une « triangulation », le triangle étant composé des deux observateurs et de l'environnement causal qui les entoure. Nous retrouvons ici l'idée que l'environnement doit agir causalement sur les interprètes. Mais comment savons-nous que ces causes sont objectives, c'est-à-dire proviennent bien d'objets du monde ? La méthode de Quine consiste à établir les relations entre les phrases observationnelles et des stimulations sensorielles. Les objets perçus sont alors seulement des « postulats » (*posits*) inférés à partir des stimulations *proximales* des sujets. C'est ce que Davidson appelle la théorie « proximale » du lien entre signification et expérience : la similarité des significations et des références est inférée de la similarité des stimuli proximaux qui déclenchent l'assentiment ou le dissentiment. Davidson ne lie pas les sti-

muli aux sensations ni aux phrases observationnelles : il les lie directement à des propriétés et à des objets *distaux* qui causent nos croyances. C'est ce qu'il appelle la « théorie *distale* de la référence » (1990c ; 1990 : 321). La cause commune de stimulations distales ne peut être identifiée qu'intersubjectivement, comme le point où se rencontrent interprète et interprété. C'est donc la communication qui rend possible l'identification des objets du monde et de ce sur quoi portent nos croyances. Mais comment nous assurer que nous avons identifié correctement les objets qui sont les causes de nos croyances ? A nouveau l'interprète et l'interprété ne peuvent-ils pas se tromper, ne peut-il y avoir communication sur base de croyances fausses ? Certes, ils le peuvent. La question, récurrente, est celle de savoir s'ils le peuvent de façon permanente et totale, au sens où le voudrait le scepticisme. Pour cela il faut préciser la nature de la relation causale qui relient les croyances à leurs causes. La relation peut être (i) une relation causale effective, (ii) une relation causale usuelle, au sens de statistiquement fréquente, (iii) une relation causale normale bien que statistiquement peu fréquente (par exemple au sens où un virus provoque normalement une maladie, bien que le virus soit rare). Il semble que la discussion ci-dessus de la similarité des réponses implique que Davidson n'entend pas parler des causes effectives, mais des causes *normales*, et que la mention des contraintes évolutionnistes lui permette de parler des causes normales au sens (iii). Ce sont des causes « subjonctives » ou contrefactuelles, au sens où c'est ce qui *causerait* une croyance en la présence d'un objet qui est la cause normale, et non ce qui le cause effectivement dans une situation donnée[1]. En ce sens, les conditions sont très voisines de celles de la théorie causale de la référence. Il ne s'agit pas de nier qu'il existe des régularités nomologiques entre les causes et les structures comportementales. Mais nous savons aussi que ces régularités ne peuvent jamais être suffisantes pour identifier les objets de référence. Le schème causal n'est jamais unique, et il y a toujours indétermination. Mais, comme

1. Cf. 1991b, 195 : « Les situations qui causent normalement une croyance déterminent les conditions dans lesquelles elle est vraie ». Ce sens peut être proche de celui qu'emploie Millikan (1984) pour parler de « fonctions normales ». Fodor et Le Pore, 1992 : 157-158, soutiennent que Davidson doit pour défendre son argument anti-sceptique soutenir que les causes sont normales au sens (i). Mais je ne vois pas pourquoi. (Ils expriment d'ailleurs leurs doutes sur ce point : 238, note 17.)

nous l'avons vu, l'indétermination n'est jamais telle qu'il ne puisse pas y avoir accord et compréhension. La « mesure du mental » n'implique pas non plus une détermination complète du mental par l'environnement causal[1]. C'est la communication seule qui est la source d'un sens et d'une vérité objective, et par conséquent d'une convergence des opinions (cf. ci-dessus, § 2.6). Davidson soutient que c'est celle-ci qui est le critère de la possession de croyances par des agents humains :

> Je soutiens que le concept d'une vérité intersubjective suffit comme base de la croyance et par conséquent pour les pensées en général. Et peut être est-il suffisamment plausible de dire que le fait d'avoir le concept de vérité intersubjective dépend de la communication au sens linguistique plein. Pour compléter l'« argument », néanmoins, je dois montrer que la seule manière par laquelle on puisse arriver au contraste subjectif-objectif passe par la possession du concept d'une vérité intersubjective. J'avoue ne pas savoir comment le montrer. Mais je n'ai pas la moindre idée de la manière dont on pourrait arriver autrement au concept d'une vérité objective. Au lieu d'un argument, j'offre l'analogie suivante.
>
> Si j'étais rivé à la terre, je n'aurais aucun moyen de déterminer la distance de nombreux objets par rapport à moi. Je saurais seulement qu'ils se trouvent sur une ligne quelconque menée d'eux à moi. Je pourrais entrer en contact avec succès avec des objets, mais je ne pourrais pas donner de contenu à la question de savoir où ils se trouvent. Comme je ne suis pas rivé au sol, je suis libre de trianguler. Notre sens de l'objectivité est la conséquence d'une autre sorte de triangulation, une qui requiert deux créatures. Chacune interagit avec un objet, mais ce qui donne à chacune le concept de la réalité objective est la ligne de base formée entre les créatures par le langage. Le fait qu'elles partagent un concept de vérité à lui seul donne un sens au fait qu'elles ont des croyances, et sont capables d'assigner à des objets une place dans un monde public (1982 : 480, 75).

L'argumentation qui se dégage de ce passage n'est pas très explicite. Nous pouvons essayer de la rendre plus explicite en nous demandant quelle est la nature de ce que l'on peut appeler l'*externalisme* de Davidson, et en quel sens cette thèse est associée à son réalisme.

1. Davidson ne souscrit cependant pas à l'externalisme défini par Putnam (1975) et Burge, pour des raisons qu'il explique dans (1987) et (1991c) et (1991e). Son externalisme est causal, mais il ne soutient pas que l'environnement, physique ou social permet à lui seul d'individualiser les contenus. Je ne discuterai pas ce point ici. Cf. Seymour, 1994a.

Dans la philosophie de l'esprit récente, on a pris l'habitude de désigner sous le nom d'*internalisme* (ou *individualisme*) la thèse selon laquelle des contenus mentaux ou psychologiques intentionnels tels que la croyance ou la connaissance de la signification d'un mot sont « internes » au sens où ils ne présupposent pas l'existence d'un autre individu que le sujet auquel ils sont attribués, et *externalisme* (ou *anti-individualisme*) la thèse selon laquelle l'individuation et la nature des contenus mentaux présupposent l'existence d'autres objets que les individus auxquels ils sont attribués. L'argument central sur lequel repose l'externalisme est l'expérience de pensée de la « Terre Jumelle » de Putnam (1975) et toutes les expériences de pensée qu'elle a inspirées, dont la forme générale est la suivante. Soient deux sujets dont les propriétés internes sont supposées identiques, mais dont les environnements extérieurs sont supposés différents. Est-il cohérent, dans ces conditions, de supposer que les états mentaux des deux sujets sont identiques ? Selon l'externalisme, ce n'est pas cohérent. Si deux sujets identiques du point de vue « interne » sont entourés d'objets et d'espèces naturelles numériquement différentes, en dépit de leur identité d'apparences externes, nous devons dire, selon l'externaliste, qu'ils ont des pensées différentes et expriment des pensées différentes par leurs mots, et par conséquent que l'individuation, et la nature, de ces pensées « dépendent » de ces objets et propriétés externes. Il y a de nombreuses variantes de ce genre d'expérience de pensée, et la nature des thèses internalistes et externalistes variera selon qu'on conçoit les états « internes » comme des états neurophysiologiques, des états fonctionnels, ou des états psychologiques spécifiques, et selon que les propriétés de l'environnement seront des propriétés physiques ou sociales. Nous n'avons pas à entrer ici dans ces distinctions. Il nous suffira, ici, de caractériser l'internalisme comme la thèse selon laquelle les états internes sont des états physiques du corps de l'agent, identifiables sans faire référence à des objets, des états ou des événements extérieurs à ce corps. La thèse externaliste qui nous intéresse est celle selon laquelle les états internes auxquels s'identifieraient les états mentaux selon l'internalisme ne peuvent pas être les états que nous nous identifions couramment quand nous parlons des croyances et des autres attitudes des agents, en particulier quand nous expliquons leur comportement, parce que l'individuation de ces états et de leurs contenus

fait typiquement référence à des objets et propriétés externes (Davidson, 1987 : 444).

Davidson rejette la thèse externaliste en ce sens (nous pouvons l'appeler externalisme *fort*). Appelons, à la suite de Putnam, « étroits » les états mentaux, quels qu'ils soient, qui ne présupposent l'existence d'autre chose que le sujet individuel auxquels on les attribue, et « larges » les états mentaux qui présupposent l'existence d'objets externes à ce sujet. Davidson admet, avec l'externalisme, que les contenus d'états psychologiques ordinaires, tels que les croyances et les désirs, sont larges en ce sens et ne peuvent pas être attribués indépendamment d'objets externes aux individus, et par conséquent que les conditions d'identité des croyances et des actions diffèrent même si les états internes restent identiques. Par exemple, on ne peut pas expliquer le comportement et les croyances que j'ai sur Terre vis-à-vis du liquide que j'appelle de l'*eau* indépendamment de l'existence d'une substance *eau* dont la composition interne est H_2O, et on ne peut expliquer le comportement et les croyances que mon jumeau physiquement identique a sur Terre Jumelle vis-à-vis du liquide qu'il appelle de l'*eau*, indépendamment de l'existence d'une substance qui est de la *jumelleau* (dont la composition interne est XYZ). Mêmes états internes étroits, même composition physique, mais différentes pensées, croyances, actions. Mais Davidson nie qu'il s'ensuive, du fait que l'on doive *décrire* et *individualiser* ces contenus mentaux en faisant référence à des objets et propriétés externes aux individus, que les états mentaux en question ne puissent pas être identifiés à des états internes *physiques* des individus. Il donne l'analogie suivante (1987 : 451). Si j'ai attrapé un coup de soleil, le fait que j'aie un coup de soleil présuppose l'existence du soleil, mais il ne s'ensuit pas que le coup de soleil ne soit pas une condition de ma peau. Parallèlement, l'identification des états mentaux par des facteurs externes aux sujets n'empêche pas que ces états soient identiques à des états physiques du corps de l'individu (et, on peut le présumer, du système nerveux central). Cette affirmation découle en fait du monisme anomal (1987 : 453) : bien que les états mentaux soient identiques à des événements physiques (neurophysiologiques) internes, leurs descriptions intentionnelles ne sont pas réductibles à des descriptions physiques. Deux êtres peuvent donc être physiquement identiques, selon Davidson, sans être

psychologiquement identiques, c'est-à-dire sans être interprétés de façon identique, et cette thèse s'acorde parfaitement avec les intuitions qui guident nos réponses aux expériences de pensée du type de celle de la Terre Jumelle. Cela n'implique pas, cependant, que Davidson accepte la distinction même entre contenu étroit et contenu large. La seule notion de contenu qu'il admet est celle de contenu large. Tout contenu est, par définition, un contenu interprétable selon la méthode d'interprétation radicale, laquelle attribue nécessairement des contenus faisant référence à des propriétés et des objets de l'environnement que l'interprète est en mesure de partager avec celui qu'il interprète. Il n'existe donc pas une autre classe de contenus susceptibles d'une individualisation étroite et correspondant aux besoins de l'explication psychologique[1].

Davidson rejette également une autre implication qu'on a parfois tirée de la thèse externaliste forte, qui est la suivante. Si le contenu de mes états mentaux est, au moins en partie, déterminé par des facteurs externes, alors cela semble devoir entraîner que je n'ai aucune autorité spéciale sur le contenu de mes propres pensées, et que je ne sais pas, en toute rigueur, ce que je signifie par mes mots (1987 : 446 sq.)[2]. Mais ce raisonnement repose sur une prémisse implicite inacceptable : que l'esprit, quand il appréhende des objets, a ces objets « devant lui », et peut les saisir comme on saisit des objets extérieurs, et que si l'identité de l'objet est déterminée par des facteurs extérieurs, alors l'identité des objets mentaux et des pensées qui portent sur eux ne peut pas être déterminée par le sujet lui-même. Or selon Davidson, il ne peut pas y avoir de tels objets « présents à l'esprit », et par conséquent la conclusion « externaliste » selon laquelle l'identité de ces objets échapperait à l'autorité de la première personne est invalide. Pourquoi ? Il convient, à nouveau, de revenir aux conditions de l'interprétation. Les pensées d'un interprète sont causalement déterminées par les objets qui l'entourent. L'interprète peut se tromper,

1. Ce point, l'unité foncière de la notion de contenu mental, et sa non-« bifurcation » en deux types selon une taxinomie proche de celle de Putnam ou une taxinomie distinguant « rôle conceptuel » et conditions de vérité, a été bien mis en avant par Bilgrami (1992). Cf. également Engel, 1992a.

2. Ici Davidson fait référence à Burge (1979) qui soutient que son anti-individualisme implique le rejet « du vieux modèle selon lequel une personne doit être directement en relation (*acquainted*), ou doit immédiatement appréhender, le contenu de ses pensées » (Davidson, 1987 : 448).

mais il aura en général raison, et en ce sens il aura une autorité sur ses pensées. Mais rien ne saurait lui garantir que ses pensées sont objectives tant qu'il ne sera pas entré en interaction avec d'autres locuteurs, et tant qu'il n'aura pas effectué la « triangulation » par laquelle il compare les contenus de ses pensées à celles d'autrui, et aux structures de stimulations qu'il reçoit. Ces conditions de l'interprétation montrent, selon Davidson, qu'il peut *à la fois* y avoir une détermination du contenu des pensées par des facteurs extérieurs (causaux) et une autorité de l'interprète sur ses pensées et les significations des mots qu'il emploie. Mais cette autorité n'est pas due à un quelconque « accès privilégié » de type cartésien à des objets présents devant l'esprit. Elle est due en dernière instance au fait que l'interprète peut être lui-même *interprété*. C'est pourquoi Davidson soutient que l'autorité de la première personne, ou la connaissance qu'un individu a à la première personne de ses contenus mentaux, est expliquée par la connaissance à la *troisième* personne, par la connaissance qu'il a des autres esprits, et par la connaissance que les autres esprits ont de lui dans des conditions communicationnelles (1984b, 1991c). Il y a bien une asymétrie entre les deux types de connaissance, mais c'est en dernier ressort la connaisance à la troisième personne, interpersonnelle, qui fonde la connaissance « personnelle » et « privée ». C'est en ce sens également que Davidson soutient qu'il est possible de parler du caractère « social » de la signification et des contenus mentaux (1992a). Revenons au « triangle » constitué par l'interprète, la personne interprétée et l'objet commun sur lequel ils doivent être capables de communiquer :

> La seule voie pour savoir que le deuxième sommet du triangle — la deuxième créature ou personne — réagit au même objet que moi, est de savoir que l'autre personne a dans son esprit les mêmes objets. Dans ce cas, la deuxième personne doit également savoir que la première personne constitue le sommet du même triangle dont un autre sommet est occupé par la deuxième personne. Deux personnes doivent être en situation de communication pour que chacun sache de l'autre qu'elles sont reliées de cette manière. Chacune d'elles doit parler à l'autre et être comprise de l'autre (1992 : 256).

L'« externalisme » de Davidson tient donc à deux conditions propres à l'interprétation. La première est l'existence d'une interaction causale entre les objets du monde et nos croyances. La seconde est le caractère

public et social des pensées et des significations dans les conditions d'une communication intersubjective. Parce que Davidson considère que ces conditions sont des conditions de possibilité de tout contenu mental et de toute signification, il considère qu'elles excluent *a priori* le scepticisme quant à l'existence du monde extérieur et le solipsisme. C'est en ce sens qu'il dit que « si l'externalisme est vrai, la question de la connaissance du monde extérieur ne se pose pas », et que « si c'est une condition constitutive de certaines pensées que leur contenu soit donnée par leur cause normale, alors la connaissance des événements et des situations qui causent ces pensées ne peuvent pas requérir qu'un sujet établisse indépendamment, ou confirme, l'hypothèse qu'il y a un monde extérieur qui cause ces pensées » (1989 : 196).

6.4. Le réalisme minimal de Davidson

Il me semble légitime d'appeler cette conception davidsonienne de l'objectivité, malgré ses dénégations, un réalisme, et de parler d'un réalisme minimal. C'est un réalisme parce qu'il maintient une distance importante entre la vérité et les croyances vraies en présupposant l'existence d'un ordre causal objectif indépendant de nous, et parce qu'il refuse d'assimiler vérité et assertabilité, y compris sous la forme de conditions d'acceptation idéales : la convergence des opinions est le fait d'interprètes finis, dont les ressources n'excèdent pas les nôtres. Ce réalisme est minimal parce qu'il n'est pas une réalisme externe, ni ne suppose une définition de la vérité comme correspondance. Enfin il est minimal parce qu'il repose sur une conception minimaliste de la vérité : le concept de vérité pris comme primitif, qui figure dans la théorie-T qui « mesure » la signification, n'a pas besoin d'être plus riche que l'équivalence de la Convention T. Mais quand la vérité est prise dans le contexte de l'interprétation, elle cesse d'être minimale en ce sens : elle se rattache à un ensemble d'intérêts, d'attitudes et de comportements humains au sein d'un monde objectif et public. En d'autres termes, la position de Davidson me paraît satisfaire au moins aux conditions (a), (b) et (e) de notre définition du réalisme minimal du § 5.5.

Savoir si une telle position peut être appelée réalisme est une chose ; savoir si elle est correcte en est une autre. Elle me paraît soumise à deux sortes de tensions caractéristiques, que nous avons déjà rencontrées.

1 / Tout d'abord, comme les autres versions du réalisme minimal que nous avons envisagées au chapitre précédent, il y a une tension, dans le réalisme de Davidson, entre d'une part la thèse selon laquelle, comme je l'ai dit au § 6.2, il ne peut pas y avoir une distance inconcevable entre nos croyances et la réalité, ou selon laquelle, en vertu du caractère « transcendantal » du réquisit de charité interprétative, il y a une sorte de limite *a priori* à l'erreur possible, et d'autre part la thèse selon laquelle la réalité a une structure causale indépendante de nous. Pour reprendre la terminologie de Wright, nous pouvons dire que c'est une tension entre l'objectivité du jugement et l'objectivité de la vérité : car même si, selon Davidson, l'interprète radical (ou son idéalisation omnisciente) peut prétendre, par un accord dans la communication, à des jugements objectifs enregistrant des propriétés d'entités réelles, il est inconcevable qu'il puisse se tromper massivement. L'intuition réaliste selon laquelle il n'y a pas de garantie que l'erreur soit impossible est donc battue en brèche. Selon la conception « interprétationniste » du réalisme minimal de Davidson, selon la conception réaliste « interne » de Putnam, ou selon la conception de la vérité comme « surassertabilité » de Wright, et quelles que soient par ailleurs les différences entre ces conceptions, la vérité est acquise par une forme d'accord ou de convergence des jugements. Mais un réaliste plus fort, ou plus traditionnel, pourra toujours insister sur le fait qu'une telle convergence ne garantit pas la vérité, et surtout qu'elle ne garantit pas l'erreur[1]. Comme je l'ai dit à la fin du chapitre 5, le réalisme minimal me paraît la position correcte dans le débat entre le réalisme et l'antiréalisme. Mais j'ai émis des doutes quant à la possibilité que la version

1. Cf. par exemple Johnston, 1991. Pettit (1991) effectue une distinction similaire à celle de Wright en assimilant le réalisme à une thèse « descriptiviste » (un discours décrit des entités distinctives), à une thèse « objectiviste » (ces entités sont indépendantes de nous), et à une thèse « cosmocentriste » (la fausseté ou l'erreur radicales sont possibles). Il soutient que tout « anthropocentrisme » selon lequel il y a une limite à l'erreur possible entre en conflit avec la thèse cosmocentriste, et assimile l'interprétationnisme de Davidson à un anthropocentrisme de ce type.

davidsonienne de cette position soit celle qui permette le mieux de dissiper cette tension. En particulier, il me semble que Davidson n'a pas montré que le scepticisme radical était impossible, et que les principes de la théorie de l'interprétation radicale suffisent à fonder une épistémologie réaliste[1]. Pour répondre à ces questions, il faudrait entrer bien plus que ne peut le faire cet ouvrage dans les questions épistémologiques et métaphysiques traditionnelles. Je me contenterai donc de considérer ici que ces questions demeurent ouvertes. La seule question que j'examinerai ici (et au chapitre suivant) est celle de savoir si le réalisme minimal de Davidson permet de satisfaire aux conditions (c) et (d) de ce réalisme énoncées au § 5.5, qui portent sur la question de l'objectivité de la *signification*. Bien que la réponse à cette question ne décide pas de la réponse aux questions de l'objectivité du jugement et de l'objectivité de la vérité, les deux réponses sont liées, dans la mesure où Davidson lie son réalisme ontologique et épistémologique à la possibilité d'assurer une théorie objective de la signification. De ce point de vue, la seconde tension que l'on peut déceler dans sa position est pertinente. C'est la suivante.

2 / Le réalisme davidsonien introduit une tension entre d'une part la thèse « causale » selon laquelle les objets de nos croyances, tels qu'ils seront identifiés par un interprète, sont les causes de nos croyances, et d'autre part la thèse « holistique » selon laquelle ce qui détermine l'objet d'une croyance ce sont les *autres* croyances que l'on peut avoir. Considérons par exemple cette déclaration de « Pensée et discours » :

> Dans quelle mesure pouvons-nous dire que les anciens — ou du moins certains anciens — croyaient que la Terre est plate ? *Cette Terre* ? Or cette Terre, la nôtre, fait partie du système solaire, lequel est pour nous identifié en partie par le fait que c'est une masse de corps célestes volumineux, froids, et solides qui tournent autour d'une étoile très grosse et très chaude. Si quelqu'un ne croit *aucune* de ces choses au sujet de la Terre, est-il bien certain que c'est de la Terre qu'il parle ? (1975 : 168, *247-248*.)

Si ce qui détermine l'objet de notre croyance ici ce sont les autres croyances que nous pouvons avoir, la possibilité d'identifier l'objet d'une croyance

1. C'est l'avis de la plupart des commentateurs de Davidson 1983 dans le volume de Le Pore, 1986 (cf. les articles de E. Sosa, C. Mc Ginn, et P. Klein).

comme étant cette Terre est indéterminée si nous ne pouvons pas attribuer de nombreuses autres croyances à propos de corps célestes volumineux, froids et solides tournant autour d'une étoile très grosse et très chaude. Mais cette indétermination est levée si nous, interprètes, faisons appel à notre croyance au sujet de la Terre, croyance qui est causée par un corps céleste volumineux et froid, et si nous appliquons le principe de charité. En ce sens interprète et interprété ne peuvent pas avoir des croyances radicalement différentes à propos d'un objet radicalement différent. Mais — et c'est ici que la tension intervient — si la relation entre les croyances et leurs contenus est une relation holistique, la causalité est une relation *non holistique* entre une croyance et un objet ou événement. Si nous admettons que ce sur quoi porte une croyance est déterminé par ce qui la cause, nous admettons aussi la possibilité que le contenu de la croyance et l'objet qui la cause ne coïncident pas. En d'autres termes, il se peut que la croyance porte sur un certain objet, mais que ce qui est cru au sujet de l'objet soit faux. La cohérence des croyances, ou leur caractère holistique, tendent à réduire la possibilité de l'erreur, alors que leur propriété d'être causées par des objets extérieurs doit admettre la possibilité de l'erreur. Il semble donc y avoir un conflit entre les deux réquisits, c'est précisément ce genre de conflit qui peut réintroduire le problème sceptique. Nous savons comment Davidson répond, en recourant à la fiction de l'interprète omniscient, et en admettant que les interprètes ne peuvent pas se tromper *massivement* sur ce qui cause leurs croyances, s'ils peuvent « trianguler » et confronter leurs croyances par le biais de la communication linguistique. Mais même dans cette dernière hypothèse, il y a un risque pour qu'une triangulation réussie autorise la communication sur des entités non existantes. C'est, semble-t-il, ce genre d'objection que ne peuvent pas manquer de soulever les partisans d'un réalisme « externe » et d'une théorie causale de la référence[1]. Leur objection est qu'il y doit y avoir une tension entre le caractère *atomique* et *naturel* de la relation de causalité entre les objets

1. Cf. par exemple Devitt, 1985. Je tire la substance de l'objection qui précède de Mc Ginn, 1977, et de Evnine, 1991 : 150-151. Mais les objections de Fodor (1987) et de Fodor et Le Pore, 1992 vont dans le même sens.

extérieurs et les croyances, et le caractère *holistique* et *normatif* du contenu des croyances dans la conception de Davidson, et que cette tension menace son réalisme.

Cette tension entre le point de vue « causal » externe et le point de vue « normatif » interne existe bien dans la philosophie de Davidson. Nous l'avons déjà rencontrée dans sa philosophie de l'esprit et de l'action, à la fois dans la thèse du caractère causal des raisons et dans le monisme anomal (§ 2.5 et 2.6)[1]. Comment, si les raisons sont des causes des actions, peuvent-elles aussi être des raisons ou des justifications de ces actions ? Comment les états mentaux peuvent-ils causer des actions *en tant que* mentaux alors que la causalité n'intervient, comme le veut le matérialisme, qu'au niveau physique ? La réponse de Davidson à ces questions est qu'il y a bien une distinction entre les descriptions physiques, causales ou « externes » d'un événement (d'une action) d'une part et les descriptions mentales, en termes de raisons, ou « internes » de cet événement ou de cette action : si cette distinction n'existait pas, le mental serait réductible au physique, et une conception naturaliste de la psychologie et de l'action humaine ne poserait aucun problème de principe. Nous devons bien être capables de voir ce même événement sous un « double aspect », l'un mental, l'autre physique, chacun de ces aspects étant hétérogène par rapport à l'autre. Mais il ne s'ensuit pas que la notion de causalité et la notion de normativité soient deux notions étrangères. Les concepts mentaux sont des concepts causaux parce qu'ils servent à expliquer les actions et autorisent des généralisations contrefactuelles semblables à celles que nous employons pour expliquer des événements physiques. Ainsi nous disons qu'un individu n'aurait pas bu un verre d'eau s'il n'avait pas eu telles croyances et tels désirs, tout comme nous disons qu'un sucre n'aurait pas fondu s'il n'avait pas été soluble. Mais si la *forme* de ces explications est causale dans les deux cas, la nature des *concepts* causaux impliqués n'est pas la même. Dans le cas des concepts causaux physiques, par exemple dispositionnels, comme la solubilité ou la malléabilité, il est possible

1. Cette tension a été notée, outre par Rorty (cf. *infra*, note 35), par de nombreux auteurs : Stoutland (1982), Follesdal (1986), Evnine 1991, et par la plupart des critiques du monisme anomal.

de réduire les explications causales à des explications nomologiques strictes dans lesquelles les notions en question auront disparu, alors qu'il n'est pas possible d'opérer une telle réduction pour les concepts mentaux. C'est en ce sens, selon Davidson que les concepts mentaux sont causaux *et* normatifs, alors que les concepts physiques ne sont *que* causaux, et peuvent, ultimement, être réduits à des concepts non causaux :

> Les propriétés normatives et causales des concepts mentaux sont reliées. Si nous devions laisser tomber l'aspect normatif des explications psychologiques, elles ne serviraient plus à ce à quoi elles servent. Nous nous intéressons si ardemment aux raisons des actions et aux autres phénomènes psychologiques que nous sommes prêts à nous contenter d'explications qui ne peuvent s'accorder parfaitement avec les lois de la physique. La physique, d'un autre côté, a pour but de découvrir des lois qui sont aussi complètes et aussi précises que possible ; c'est un objectif différent. L'élément causal dans les concepts mentaux nous aide à remplacer la précision dont ils manquent ; cela fait partie du concept d'une action intentionnelle qu'elle soit causée et expliquée par des désirs et des croyances ; cela fait partie du concept de croyance et de désir que croyances et désirs tendent à causer, et ainsi à expliquer, des actions d'un certain type (1991b : 163).

En d'autres termes, Davidson renverse ici l'association usuelle du « causal » au seul « physique » : c'est *parce que* les concepts mentaux sont irréductibles à des concepts physiques qu'ils sont « causaux », et c'est parce que les concepts physiques sont réductibles à des concepts ultimement non causaux qu'ils ne sont pas ultimement causaux. Il en est de même des concepts relatifs à la théorie de l'interprétation et du langage : il n'y a pas d'incompatibilité entre le caractère intrinsèquement « sémantique » et « normatif » des descriptions que donne un interprète du comportement linguistique et le fait que l'interprétation implique que l'on envisage la relation causale entre les sujets et leurs environnements. Du point de vue de Davidson par conséquent, la tension entre le point de vue de l'interprète qui l'amène à favoriser les considérations holistiques et de rationalité propres aux concepts mentaux, et le point de vue causal impliqué par son rejet de l'empirisme et par l'idée que les contenus mentaux sont déterminés par le monde sans intermédiaires « épistémiques », n'existe pas.

Elle n'existe pas parce que sa théorie de l'interprétation l'amène à traiter ces considérations sur le même plan[1].

Cette réponse de Davidson ne peut satisfaire que si l'on admet les principes du monisme anomal et de sa théorie du mental et de l'action. Je n'ai pas prétendu les justifier ici, et, par conséquent je laisserai également cette question ouverte. Mais si l'on met à part le problème de la compatibilité entre le point de vue causal et le point de vue normatif, ce que la tension en question révèle est la nécessité de réexaminer les principes de son holisme.

1. *Petite digression rortyenne*. Cette réponse permet aussi d'écarter certaines critiques adressées à Davidson par Rorty, mais aussi certaines associations effectuées par ce dernier entre ses thèses et celles de Davidson. Rorty entend nier que Davidson soit un réaliste, et le féliciter de ne pas en être un. Il cherche à exploiter l'argumentation de Davidson, 1974. Selon lui, si nous prenons au pied de la lettre la critique par Davidson du dualisme du schème et du contenu, et si nous admettons que la vérité d'un énoncé dépend seulement de deux choses — ce que les mots signifient et les choses qui existent dans le monde — il n'y a plus lieu de parler de correspondance, et par conséquent plus lieu de parler de réalisme (1986 : 309). Rorty pense que la position de Davidson rend vide la question de savoir si nous pouvons avoir une connaissance objective d'un monde public indépendant de nous. Il soutient que le rejet de toute théorie de la vérité comme correspondance implique qu'il n'y a aucun lien entre la question de savoir quelles relations inférentielles existent entre nos croyances (l'exigence holistique) et la question de savoir sur quoi portent nos croyances, et ce qui les cause (l'exigence causale). La seconde question requiert que nous nous placions d'un point de vue externe et causal, alors que la première requiert que nous nous placions d'un point de vue interne, portant sur la justification de nos assertions. Le premier point de vue est normatif, le second descriptif. Or il n'y a aucun moyen, selon Rorty, de concilier ces deux points de vue. Ils racontent simplement deux « histoires » distinctes, et recouvrent deux attitudes différentes mais irréconciliables que nous pouvons adopter sur nos « pratiques linguistiques ». Selon Rorty, la thèse principale de Davidson est que ces deux points de vue doivent être séparés, et correspondent à deux sens distincts du terme « vrai ». Selon le point de vue « normatif », « vrai » n'est qu'un terme « appréciatif » — « un compliment que nous adressons à nos assertions » — alors que selon le point de vue « causal », « vrai » rapporte le fait que les êtres humains sont causalement affectés par des choses du monde. D'après Rorty, par conséquent, la position que devrait adopter Davidson est celle d'un « pragmatiste », et il définit ce « pragmatisme » comme la conjonction de thèses suivantes :

(1) « Vrai » n'a pas d'usage explicatif.

(2) Nous comprenons tout ce qu'il y a à comprendre quant aux relations entre nos croyances et le monde quand nous comprenons leurs relations causales avec le monde ; notre connaissance de la façon dont nous appliquons des termes tels que « est à propos de » ou « est vrai de » découle d'une analyse « naturaliste » du comportement humain.

(3) Il n'y a pas de relations d'« être rendu vrai par » entre nos croyances et le monde.

(4) Les débats entre le réalisme et l'antiréalisme n'ont pas de sens, parce que ces débats présupposent l'idée vide et trompeuse des croyances « rendues vraies » (Rorty, 1986 : 335).

Je ne soulèverai pas ici la question complexe de savoir si Rorty emploie le terme « pragmatisme » à bon escient, bien qu'on puisse en douter, dans la mesure où ces thèses sont supposées s'appliquer

6.5. Du bon usage du holisme

Nous avons pu, à de nombreuses reprises, voir en quoi la philosophie du langage et de l'esprit de Davidson reposait sur une certaine forme, ou certaines formes, de holisme. Nous avons également défendu ce holisme contre les objections qu'on lui a adressées (§ 2.6, § 4.3, § 4.5, § 4.6). Il est temps d'essayer d'envisager cette notion pour elle-même et de voir où nous en sommes sur ce point. Cet examen est d'autant plus nécessaire que nous avons qualifié la position de Davidson de réaliste, et que la plupart de ses critiques soutiennent que le holisme menace le réalisme.

(Suite de la note 1, p. 262.)
aussi, outre à Rorty lui-même, à James, à Dewey, à Nietzsche, à Heidegger, à Derrida et à Foucault (*sic*) mais pas à des auteurs comme Peirce. Mais il est indéniable que les thèses (1)-(4) peuvent être attribuées à Davidson, tout comme elles caractérisent une certaine forme de déflationnisme ou de minimalisme quant à la vérité, au sens analysé au chapitre précédent. Mais nous avons précisément rejeté l'idée que le réalisme de Davidson soit un déflationnisme en ce sens. Par conséquent (1)-(4) ne peuvent pas caractériser la position davidsonienne, et ce dernier ne peut pas être un « pragmatiste » au sens où l'entend Rorty. Cette question est cependant dans une large mesure terminologique, et on la laissera de côté. La question importante est celle de savoir si Davidson entretient bien l'idée d'une séparation radicale entre le point de vue « normatif » interne et le point de vue « causal » externe, et si cette séparation menace, comme le soutient Rorty, son réalisme.

Rorty (*ibid.* : 353) suggère qu'une théorie qui s'efforce de combiner le point de vue causal sur nos croyances et sur nos actions et le point de vue de leur justification commet le type d'erreur que Wittgenstein nous a appris à ne pas commettre : confondre la vérité d'une croyance avec une relation réelle (causale) à une entité du monde, et confondre les raisons et les causes. Il soutient que les deux points de vue sont incompatibles, et que Davidson a défendu leur incompatibilité. Davidson au contraire défend leur compatibilité. Il ne s'ensuit pas qu'il commette l'erreur, dénoncée par Wittgenstein, de *confondre* les deux points de vue, ni qu'il n'y ait aucune manière de montrer leur compatibilité. Toute la philosophie de l'action de Davidson vise à montrer comment les raisons peuvent être des « causes rationnelles » et comment il est possible d'adopter un point de vue à partir duquel les raisons d'agir peuvent être conçues *simultanément* comme justifiant les actions et comme les causant. En philosophie du langage, Davidson soutient que l'interprétation repose sur des conditions causales, bien que ses produits soient sémantiques, et propres à figurer dans des justifications. Nos croyances sont à la fois des entités porteuses d'un contenu, soumises à des conditions normatives de rationalité, et reliées causalement au monde. Si la position que j'ai appelée « réalisme minimal » est cohérente, il doit être possible de reconnaître à la fois que la notion de vérité a un sens descriptif et qu'elle a un sens normatif (celui de la *Vérité*) sans que l'un des sens se réduise à l'autre. Rorty au contraire entend donner à la notion de vérité un sens uniquement normatif, ou évaluatif, celui d'une simple approbation de nos assertions, sans aucun lien avec une quelconque réalité. La vérité est seulement la propriété qu'attribuent à leurs assertions les membres d'une communauté ou ceux qui partagent un type de discours, et jamais une propriété qui pourrait

Nous avons rencontré, jusqu'à présent, plusieurs formes de holisme. (i) La première est ce que nous avons appelé (§ 1.2) le *holisme de la phrase*, correspondant au principe fregéen de contextualité (où plutôt à une certaine interprétation de celui-ci) : ce n'est que dans le contexte d'une phrase que les mots ont une signification, et la phrase complète est le premier porteur des propriétés et valeurs sémantiques. (ii) La seconde est ce que nous avons appelé le *holisme du langage* ou *holisme de la signification* : la signification des expressions simples et des phrases d'une langue dépend de la signification d'autres, voire de toutes les expressions et phrases de cette langue. (iii) La troisième est le *holisme épistémologique*, ou *holisme de la confirmation* selon lequel une phrase ne peut jamais être confrontée à l'expérience isolément, parce que seuls des ensembles de phrases rencontrent le tribunal de l'expérience. Ces trois formes principales du holisme devraient encore être diversifiées selon qu'elles portent sur des *phrases* et sur leurs significations (holisme proprement *sémantique*), ou selon qu'elles portent sur des *croyances* ou des contenus (holisme *du mental*) ou *holisme psychologique*). Nous avons admis, jusqu'à présent, que Davidson souscrivait à ces trois formes principales du holisme. Mais il souscrit également à une quatrième forme, plus générale encore, de holisme, que l'on peut appeler *holisme de l'interprétation* : parce que, dans l'interprétation

(Suite et fin de la note 1, p. 262)
s'attacher à des énoncés en vertu d'une quelconque correspondance avec la réalité. Davidson est, comme l'on a vu, d'accord sur le second point. Mais il nie que du fait que la vérité soit une norme de la correction des assertions, il s'ensuive qu'il n'y a aucune forme d'objectivité que puisse atteindre nos jugements, comme nous l'avons vu avec les diverses versions du réalisme minimal envisagées au chapitre précédent. Rorty au contraire soutient qu'on ne peut pas parler d'une quelconque objectivité et que toute forme de savoir doit se ramener à une forme de pratique sociale sur laquelle s'accordent ses participants, et que le mieux que l'on puisse faire est de « maintenir la conversation » entre les participants de diverses pratiques, sans qu'ils puissent jamais parvenir à une quelconque convergence d'opinions. Il n'y a plus de place pour une théorie de la connaissance, seulement pour un projet « herméneutique » (Rorty, 1979). Bien que Rorty prétende tirer ces conséquences des critiques de l'épistémologie classique « fondationnaliste » avancées par Quine et Davidson eux-mêmes, il n'y a qu'une ressemblance superficielle entre sa position et celle de Davidson. J'espère avoir montré qu'adopter le point de vue de l'interprétation au sens où ce dernier l'entend n'implique ni un renoncement à l'idée qu'on puisse avoir une connaissance d'un monde objectif, ni une réduction de la vérité à « un compliment » que nous adressons à nos assertions. On voit mal alors en quoi le prétendu « pragmatisme » que Rorty attribue à Davidson (et le sien propre, s'il entend défendre cette position) pourrait se distinguer des formes de relativisme et d'antiréalisme que Davidson critique. Cf. aussi Haack 1993.

radicale, il est impossible de séparer les significations, les croyances, les désirs, et les actions des individus, on ne peut interpréter les unes sans interpréter les autres, et on ne peut interpréter un contenu sémantique ou psychologique sans le rattacher à d'autres contenus. Ce holisme de l'interprétation est étroitement associé au principe de *l'interdépendance des croyances et des significations*, et il conduit Davidson à considérer comme indissociables le holisme sémantique et le holisme psychologique. C'est ce holisme que l'on a vu à l'œuvre dans la théorie généralisée de la signification et de l'action (§ 2.7), et c'est celui que caractérise McDowell quand il parle de « l'enchevêtrement » des contenus de phrases, des actes linguistiques, et des états psychologiques des membres d'une communauté (§ 4.6). Il est frappant de constater, comme nous l'avons déjà fait, que ce holisme de l'interprétation est, chez Davidson, essentiellement *méthodologique*, et qu'il l'invoque le plus souvent à l'appui des autres formes de holisme : c'est parce que les conditions dans lesquelles nous sommes conduits à *attribuer* à des individus divers contenus sémantiques ou psychologiques sont holistiques que nous pouvons en inférer que les significations, et les contenus mentaux, le sont. Et c'est ce raisonnement qui fait problème pour les critiques de l'« interprétationnisme » de Davidson : est-on en droit d'inférer, à partir des conditions de vérification et d'interprétation des contenus, quelque chose portant sur leur *nature* ? Peut-on passer d'un holisme *méthodologique* à un holisme *substantiel* ou *métaphysique* ? Davidson semble ici pris dans un dilemme. Ou il accepte ce raisonnement vérificationniste, mais il doit alors concéder qu'il défend une conception purement antiréaliste des contenus (ceux-ci ne sont pas réels, mais attribués ou interprétés). Ou il ne l'accepte pas, et il doit admettre qu'il n'a pas donné d'argument probant, sur la base des considérations méthodologiques de l'interprétation, quant à la nature même des contenus. Je me suis efforcé, à plusieurs reprises (§ 2.6 ; § 4.6), de montrer que la conception davidsonienne échappe à ce dilemme. Mais je voudrais ici tempérer mes analyses précédentes.

Deux sortes de problèmes se posent à nous. Le premier est celui des conséquences catastrophiques que sont supposées avoir les diverses thèses holistes pour la nature de la signification, du langage, et des contenus mentaux. Le second est celui des liens logiques qu'entretiennent entre

elles les diverses thèses. Ces deux problèmes sont liés. Quelles sont ces conséquences alléguées ? Ce sont, comme nous l'avons vu, (a) que le holisme de la signification interdit tout apprentissage du langage et toute maîtrise possible, par un locuteur, des significations de ses mots, que (b) il interdit *toute* théorie de la signification satisfaisant le réquisit de manifestabilité du sens, (c) qu'il interdit de dire jamais qu'une phrase ait une signification (Dummett). Quant au holisme psychologique, on prétend qu'il interdit de dire (a) qu'un individu ait une croyance déterminée, (b) que deux individus aient la même croyance, et (c) de faire la moindre généralisation sur les croyances et autres états mentaux des individus et par conséquent de formuler des lois intentionnelles (Fodor). On doit admettre que ces conséquences sont désastreuses, en particulier pour la possibilité même d'établir une théorie systématique de la signification, qui nous a occupé tout au long de ce livre. Mais je me suis efforcé de montrer que le holisme de Davidson n'avait pas ces conséquences, et que leur rejet ne justifiait pas que l'on adopte, comme Dummett, une conception antiréaliste de la signification, ou, comme Fodor, une conception hyperréaliste. Ce qui suit est une autre tentative pour écarter ces critiques anti-holistes, tout en reconnaissant, dans une large mesure, leur légitimité.

La cible principale des critiques est respectivement le holisme de la signification ou du langage (le sens d'un mot ou d'une phrase dépend du sens d'autres mots ou phrases, voire du sens de *tous* les mots ou phrases du langage) et le holisme psychologique (le contenu d'une croyance d'un individu dépend du contenu d'autres, voire de *toutes* les croyances de cet individu). C'est précisément parce qu'il semble impossible, si l'on admet ces thèses, de jamais pouvoir isoler un contenu sémantique ou psychologique déterminé, que les adversaires du holisme les rejettent, soit que, comme Dummett, ils y voient l'effet d'un réalisme outrancier (la signification échappe à toute recognition), soit que, comme Fodor, ils y voient l'effet d'un antiréalisme outrancier (la signification est indéterminée par essence). La plupart des critiques du holisme de la signification et du holisme des croyances contestent donc essentiellement la thèse (ii), et contestent les thèses épistémologiques (iii) et (iv) dans la mesure où ces thèses sont tenues comme justifiant (ii). Supposons, pour le moment, que la thèse (ii) soit bien celle de Davidson, et qu'elle ait les conséquences

désastreuses qu'elle est supposée avoir. Quelle sorte d'argument peut la justifier et cet argument est-il valide ? Il y a, semble-t-il, deux sortes d'arguments possibles, qui partent chacun de prémisses plus faibles que (ii).

(1) Le *premier argument* part d'une prémisse en apparence incontestable, le principe de compositionnalité, selon lequel le sens d'une phrase est déterminé par le sens des expressions qui la composent, et du principe (i), le holisme de la phrase. Admettons, selon ce principe, que l'unité minimale de la signification soit la phrase. Soit une phrase quelconque. Par compositionnalité, son sens dépend du sens des expressions qui la composent. Le sens de ces dernières dépend du sens de toutes les phrases dans lesquelles elles figurent, et le sens de celles-ci dépend en retour du sens de leurs constituants, et ainsi de suite pour tout le langage. C'est apparemment un raisonnement de ce type qu'emploie Davidson dans le passage (1967 : 22, *48*) cité au § 1.1. Qu'est-ce qui le justifie ? La principale justification est celle que Davidson fournit dans sa critique de la théorie de l'empilage des blocs (§ 6.1, ci-dessus). Il rejette cette théorie, comme on l'a vu, pour une raison essentiellement épistémologique : ce n'est qu'au niveau des phrases qu'une théorie de la signification peut être confirmée par des données non linguistiques, et en ce sens son holisme de la signification et des croyances repose sur le holisme (iv) de l'interprétation. Pour un critique du holisme, il n'y a, semble-t-il, que trois manières de rejeter l'argument (1). Si ce critique souscrit au principe de compositionnalité, il peut : (a) rejeter le holisme de la phrase qui est l'une des prémisses de cet argument, tout en acceptant le holisme méthodologique de l'interprétation ; (b) rejeter le second et accepter le premier ; et (c) rejeter ces deux holismes. Chacune de ces options entraîne une certaine forme d'atomisme ou de molécularisme. La première option (a) n'est pas très attrayante. Elle suppose, conformément à l'atomisme et à la théorie des blocs empilés, que le sens des expressions d'un langage puisse être fixé indépendamment de celui des phrases, tout en admettant que c'est au niveau des phrases, et selon une méthode holistique, que l'on teste une théorie de la signification. On voit mal alors en quoi ces deux propositions ne peuvent pas être contradictoires. Quant à (b), elle admet la primauté de la phrase sur les mots, mais nie que cela ait la moindre conséquence sur la manière dont on teste la signification des phrases. A nouveau, cela apparaît peu plausible.

La voie la plus plausible, pour l'antiholiste, semble être donc (c). Il doit pouvoir soutenir que le sens des expressions simples d'un langage est *par nature* indépendant du sens des phrases dans lesquelles elles figurent, et que le sens des premières est *déterminé indépendamment* du sens des secondes. Or cela suppose, pour toute version atomiste de cette option, soit qu'on ait des moyens de déterminer les sens des expressions isolés dans le comportement observable indépendamment des attitudes des locuteurs vis-à-vis de phrases et de leur vérité, soit qu'il existe des moyens, *indépendants* du comportement des locuteurs, de fixer ces significations. La première possibilité correspondrait, par exemple, à une théorie de l'apprentissage des termes, qui impliquerait qu'on puisse déterminer le sens, par exemple, des noms propres d'un langage comme termes servant à nommer des objets, sans que l'on ait besoin de passer par les attitudes des locuteurs vis-à-vis de phrases les contenant. La seconde possibilité correspondrait à une forme de théorie causale de la référence du type de celles envisagées au § 6.1 ci-dessus. Je n'exclus pas que l'une ou l'autre, ou une combinaison de ces deux théories, puisse être établie[1]. Mais elles se heurteraient à l'argument dummettien de la *manifestation*, ou à sa version davidsonienne. La première solution, si elle admet que le sens d'expressions isolées peut être établi à partir du comportement et des usages des locuteurs, doit montrer qu'il est possible de le faire sans les contraintes holistiques de l'interprétation du comportement. Mais si l'on admet, comme nous l'avons fait avec Wright (§ 5.2), que c'est « entre l'aptitude [à utiliser une expression] et sa manifestation que tombe l'ombre du holisme », toute théorie atomiste et molécullariste, qu'elle soit antiréaliste ou réaliste, doit faire la preuve que les contraintes de la manifestation peuvent être satisfaites *non holistiquement*. Et c'est précisément ce qu'il apparaît impossible de faire[2]. Par conséquent l'atomiste et le molécullariste doivent ou bien trouver

1. Par exemple, pour une version de la première théorie, Panaccio (1992) soutient qu'une théorie « occamiste » des termes serait suffisante. Pour une version de la seconde, on peut envisager la théorie causale de Fodor (1987, 1992).

2. Ainsi Panaccio (*ibid.*) dans son intéressante critique du holisme de Davidson (194-205), propose une théorie occamiste des blocs empilés, d'après laquelle « on pourrait identifier des types de comportements typiques liés à la nomination (ou à la connotation) » et soutient que répondre que ces comportements présupposent l'usage de phrases est « une pétition de principe ». Mais d'une part il ne dit pas comment on pourrait, dans des comportements linguistiques, opérer les identifications

une version du réquisit de manifestation qui ne soit pas holistique — et c'est la difficulté à laquelle, comme nous l'avons vu, se heurte la version dummettienne de la manifestation — soit rejeter purement et simplement ce réquisit. C'est, me semble-t-il, le sens du défi qu'adresse Davidson à toute forme de théorie des « blocs empilés ».

(2) On peut envisager un *second argument* en faveur du holisme de la signification et des croyances, celui que reconstruit, pour le critiquer, Fodor sous le nom d'*Ur-Argument* (1987) et d'argument-A (Fodor et Le Pore, 1992), qu'on peut formuler ainsi[1] :

> (a) On définit d'abord une propriété holistique en général à partir de la notion de propriété *anatomique* : Une propriété P est anatomique si et seulement si une chose la possède, alors de nombreuses autres choses la possèdent. Une propriété est holistique si elle est « très anatomique », c'est-à-dire si, si une chose l'a, alors de *nombreuses* choses l'ont (1992 : 2). Il définit aussi une propriété *sémantique* comme la propriété pour une phrase d'avoir une certaine signification, et pour une croyance un certain contenu.
> (*Prémisse 1*) Des propriétés sémantiques génériques, telles que être une croyance, ou être une phrase d'un langage sont anatomiques, *i.e* si un individu a une croyance *p* (si une phrase exprime la signification que *p*) alors il doit avoir un grand nombre d'autres croyances non identiques à *p* (il doit y avoir de nombreuses autres phrases exprimant des significations non identiques).
> (*Prémisse 2*) Il n'y a pas de distinction justifiée entre les propositions que l'on doit croire pour croire que *p* et les propositions que l'on ne doit pas croire pour croire que *p* (et pas de distinction entre les significations que doit exprimer la phrase *p*) et celles que cette phrase ne doit pas exprimer.
> (*Conclusion*) La propriété d'être une croyance (ou d'être une signification) est holistique (Fodor et Le Pore, 1992 : 23-25).

Selon Fodor et Le Pore, le poids de cet argument repose à la fois sur la prémisse *1*, l'anatomisme des croyances et des significations, et sur la

(Suite de la note 2, p. 268)
en question, et d'autre part il fait lui-même une pétition de principe contre le holisme méthodologique, quand il soutient que ces comportements pourraient être identifiés indépendamment des autres croyances et attitudes des locuteurs. L'« ombre du holisme » tombe aussi sur une méthode de manifestation de capacités à employer des types d'expressions isolées.

1. Les deux arguments ne sont pas identiques, et la formulation qui suit n'est pas exactement celle de Fodor, qui s'adresse surtout au holisme des croyances, mais rien d'essentiel pour le présent propos ne découle des modifications que je fais ici.

prémisse 2, qui repose elle-même sur le rejet de la distinction analytique/synthétique, et sur une forme de raisonnement sorite (*slippery slope*), selon lequel si, pour avoir une croyance, on doit avoir d'autres croyances, rien n'interdit qu'on ait un nombre *indéfini* d'autres croyances — d'où le holisme. Ils ne voient rien à objecter en soi à la critique de la distinction analytique/synthétique, ni à l'emploi d'un tel raisonnement sorite. Je n'entrerai pas ici dans l'examen de leurs raisons pour ces affirmations, puisqu'ils entendent faire porter leur critique sur la prémisse 1. Ils soutiennent (1992 : 28) que (1) contient une ambiguïté de portée, et peut vouloir dire soit que (a) *Il y a d'autres propositions telles que l'on ne peut croire p sans les croire* (portée large, anatomisme *fort*), soit que (b) *On ne peut croire p s'il n'y a pas d'autres propositions que l'on croit* (portée courte, anatomisme *faible*) et que seule cette seconde version entraîne le holisme de la signification. Mais comme le fait remarquer Laurier (1994) dans sa discussion détaillée de ce point, il n'est pas évident que le holisme de Davidson soit commis à plus que (b), la version faible. C'est en particulier ainsi, semble-t-il, qu'on peut comprendre son raisonnement cité ci-dessus (§ 2.3) au sujet de la croyance qu'un nuage passe devant le soleil. En ce cas, son holisme serait beaucoup plus faible que ne le soutiennent Fodor et Le Pore. Mais admettons que ce soit bien la version forte qui corresponde au holisme de Davidson. L'idée à laquelle Fodor et Le Pore s'opposent est l'idée même selon laquelle les propriétés sémantiques (significations ou contenus mentaux) seraient « anatomiques », et la thèse qu'ils veulent défendre est l'atomisme sémantique (une phrase peut exprimer une signification isolément) et psychologique (un individu peut n'avoir qu'une seule croyance, être un esprit « ponctué »). Or il est clair que si le holisme auquel souscrit Davidson repose bien sur la prémisse 1 au sens où l'interprètent Fodor et Le Pore, les arguments qu'il emploie pour défendre ce holisme s'appuient tous, une fois encore, sur la méthodologie de l'interprétation. Comme on l'a vu au § 2.3, il soutient que les conditions d'attribution des croyances et des significations, et l'usage des principes normatifs de rationalité, comme le principe de charité, nous forcent à attribuer un *ensemble* de significations et de contenus mentaux à des locuteurs. Fodor et le Pore le voient bien, quand ils concentrent leurs critiques (1992, chap. 3) sur la théorie davidsonienne de l'interprétation. Ils entendent montrer que les principes de cette théorie ne permettent pas de choisir entre diverses théories extensionnellement équivalentes de la signification proposées par un interprète, et qu'ils n'impliquent pas le holisme de la signification sous sa forme métaphysique portant sur la *nature* des significations et des croyances. La première critique n'est pas différente des diverses critiques (de Foster, de Dummett) que nous avons examinées précédemment. Et Davidson n'a rien à y redire. Il soutient précisément que sa méthode d'interprétation ne garantira pas l'unicité des théories de la signification, bien que sa méthode permette de réduire l'indétermination. Par conséquent si l'atomiste ou le moléculariste veulent défendre ici l'idée, une autre méthode de fixation du sens pourra le déterminer complètement, la charge de la preuve leur incombe. La seconde critique seule a une portée véritable, s'il est vrai qu'elle a les conséquences désastreuses qu'est supposé avoir le holisme de la signification.

La réponse davidsonienne à cette critique, que nous avons déjà donnée ci-dessus au sujet des objections de Field (§ 6.1), est prévisible. Considérons, une fois de plus, la situation de l'interprète radical (dans une interprétation individuelle). Il repère un ensemble de phrases tenues pour vraies par un locuteur, et se met, sur la base de ses propres croyances et des données environnementales dont il dispose et de ses assignations charitables, à attribuer des significations à des phrases du langage interprété et des croyances au locuteur. A *ce* stade, rien ne semble devoir limiter le holisme. L'interprète est contraint de faire des assignations sur un ensemble de phrases, et de vérifier la récurrence de divers types d'expression sur cet ensemble. Il ne peut dissocier nettement les significations des phrases des croyances (et désirs) qu'entretient le locuteur, et ne peut attribuer l'une d'elles sans en attribuer un certain nombre d'autres. Mais une fois que l'interprète sera parvenu à une interprétation stable du discours du locuteur, il sera en mesure de désimpliquer les rôles respectifs des croyances (désirs) et des significations, et de séparer les croyances (et préférences) effectivement tenues pour vraies par le locuteur de celles qu'il avait seulement attribuées sur une base charitable. On peut ainsi effectuer une distinction comparable à celle, énoncée au § 6.1, entre deux niveaux, celui de l'explication *interne* à la théorie et celui de l'explication *de* la théorie : *dans* la théorie proposée par l'interprète, les ensembles de croyances

et de significations ne sont pas correctement délimités, et le holisme règne, alors que dans l'explication *de* la théorie l'interprète est en mesure d'assigner une structure plus fine à ces croyances et significations.

J'ai soutenu, dans ce qui précède, notamment contre Dummett, que le holisme de Davidson était avant tout méthodologique, qu'il n'était pas évident qu'il soit « constitutif » au sens de la métaphore du « réseau », et qu'il était, comme l'a vu Tennant, compatible avec le molécularisme d'une théorie-T. Il n'y a aucune raison de supposer, en particulier, que la procédure d'interprétation radicale entraîne que la signification d'une expression ou d'une phrase, ou le contenu d'une croyance, dépendent de la signification ou du contenu de *toutes* les autres phrases et croyances d'un locuteur. Rien n'indique, de ce point de vue, que le holisme de Davidson soit plus fort que la thèse (b) ci-dessus selon laquelle on ne peut croire que *p* sans croire d'*autres* propositions distinctes de *p*. En particulier si la méthode d'assignation simultanée de croyances, de désirs, et de significations décrite au § 2.7 s'applique, l'interprète doit pouvoir déterminer les facteurs interdépendants de son équation initiale, et il doit pouvoir, par cette méthode de « mesure » du mental, parvenir à des assignations *suffisamment* uniques.

Mais les critiques du holisme ne sont pas satisfaits par cette réponse, pour deux raisons. Tout d'abord, ils font remarquer, comme Dummett ou Fodor, que si le holisme fonctionne au premier niveau, dans la théorie, sur le plan méthodologique, on voit mal ce qui peut le limiter au second niveau, qui applique la théorie, et par conséquent comment on pourra jamais parvenir à des assignations de significations, de références, et de croyances stables et définitives. Ensuite, ils feront remarquer qu'il n'est pas évident que le holisme soit seulement méthodologique même au premier niveau. Si Davidson soutient qu'il n'est *jamais* possible de spécifier ce que quelqu'un signifie sans spécifier ce qu'il croit, y compris aux stades ultimes (ou provisoirement ultimes) de l'interprétation, alors il défend bien une forme de holisme métaphysique quant à la nature des contenus. Et s'il soutient que le principe de charité implique qu'aucune phrase-T ne peut jamais être confirmée tant que l'ensemble de la théorie-T ne l'a pas été, alors il défend bien un holisme radical non seulement quant à la confirmation d'une théorie de la signification mais quant à la nature

des significations elles-mêmes, s'il défend également l'idée que l'interprétation constitue la signification[1].

On doit admettre que le holisme de Davidson doit bien être métaphysique plutôt que seulement méthodologique. Si Davidson n'entendait pas, à partir de sa conception de la procédure d'interprétation radicale, tirer des conclusions sur la nature des significations et des croyances, son réalisme des « trames » (§ 2.6) et ce que j'ai appelé son réalisme minimal n'aurait pas de sens. Davidson soutient bien qu'une expression ou une phrase n'a de sens, et une croyance de contenu, *que* si elle est radicalement interprétable. Et puisqu'il n'existe pas, dans la procédure interprétative, de principes qui permettent de limiter l'application du principe de charité, ou que, si de tels principes existent, ils relèvent purement de la psychologie, que Davidson laisse indéterminée, il faut admettre que le holisme davidsonien doit être de l'espèce radicale. Bien que j'aie défendu ce holisme contre ses diverses critiques, je concéderai la force de ces critiques — tout comme j'ai accepté certains principes antiréalistes au chapitre précédent. Mais je n'en conclurai pas que des théories atomistes ou molécularistes de la signification soient pour autant correctes. Il nous faut donc proposer un autre cadre d'analyse.

Le cadre que je proposerai est très similaire à celui qui nous a conduit à formuler le réalisme minimal au chapitre précédent. Tout comme on peut formuler une conception minimale de la vérité, réduite aux platitudes associant vérité et assertion, nous pouvons formuler un *holisme minimal*, réduit aux platitudes suivantes :

> *(a) Première platitude : le holisme des attributions de croyances et de significations.* Aucun comportement ne peut être considéré comme distinctif de la connaissance qu'a un locuteur de la signification d'une expression ou d'une phrase, ni comme distinctif d'une croyance particulière. Un seul type de comportement peut être relié à de nombreuses significations, être l'expression de nombreuses croyances, et une signification ou une croyance peut trouver son expression dans de nombreuses sortes de comportement.
> *(b) Deuxième platitude : compositionnalité et systématicité.* La signification d'expressions linguistiques complexes est fonction des significations d'expressions simples, et le contenu des pensées est fonction des concepts qui les composent. Il y a des liens systématiques entre les phrases et les pensées, en particulier des liens inférentiels.

1. Comme le remarque Laurier, 1994.

(c) *Troisième platitude: généralité et productivité*. Si quelqu'un connaît la signification de *a est F*, et s'il connaît le sens d'un nom *b*, alors on peut présumer qu'il connaît la signification de *b est F*, et de nombreuses autres phrases de cette forme composées d'autres noms propres (et pareillement pour des prédicats distincts de *F*).

J'appelle ces trois principes des platitudes parce qu'ils paraissent évidents et incontestables, et parce que les critiques du holisme eux-mêmes les admettent. Considérons (*a*). Peut-on nier sérieusement, sans souscrire à une forme rudimentaire de béhaviorisme qui associerait la maîtrise d'une signification ou la possession d'une croyance à *un seul* type de comportement, qu'une croyance (ou une signification) qui conduit typiquement à un certain type de comportement peut trouver son expression dans un autre, ou dans de nombreux autres comportements, y compris contraires ? Je peux manifester ma croyance que la mer est froide en refusant de me baigner, mais aussi en plongeant vaillamment, si je désire aussi montrer mon courage. Une croyance n'est associée à un comportement que moyennant d'autres croyances, lesquelles peuvent être modifiées par d'autres croyances ou états mentaux, tels que des désirs. De même un comportement linguistique peut être l'expression de nombreuses croyances et états mentaux distincts. Il semble difficile de nier *cette* forme de holisme, au moins quand elle ne porte que sur la vérification des croyances et des significations. On peut soutenir que Davidson ne fait pas appel à un principe différent, quand il souligne la difficulté d'attribuer des croyances indépendamment d'autres croyances ou états mentaux, et les significations indépendamment des croyances[1]. Le holisme méthodologique est donc tout à fait justifié. Considérons maintenant (*b*), la compositionnalité. Elle est admise par la plupart des linguistes, bien qu'on puisse la contester dans certaines conditions[2]. Certains auteurs comme Schiffer (1987 : 212 sq.) l'ont contestée plus systématiquement. Mais d'une part il n'est pas évident que ces critiques soient concluantes, et d'autre part personne ne nie qu'il s'agisse d'un principe de prime abord évident, qui mérite en ce sens le titre de

1. Comme le remarque Laurier, 1994.

2. Il recourt précisément à ce principe quand il critique le « béhaviorisme définitionnel » (1970 : 217). Cf. également Wright, 1987 : 22, auquel j'ai emprunté l'exemple ci-dessus.

platitude[1]. Considérons enfin (*c*), la capacité pour un locuteur de former et de comprendre, à partir de phrases contenant des expressions simples, des phrases distinctes de même structure contenant d'autres expressions du même type. Comme la compositionnalité, cela paraît être une capacité linguistique et conceptuelle élémentaire, qui fait partie de la productivité même du sens linguistique et de la pensée. Il ne s'agit pas de nier qu'il y ait des contre-exemples possibles à ces trois platitudes : certains locuteurs peuvent être si peu sophistiqués qu'ils ne sont capables d'associer qu'un seul comportement à une phrase ou à une croyance ; certains sujets peuvent avoir des lésions affectant leur faculté linguistique telles que leurs capacités à généraliser ou à inférer systématiquement soit bloquée. Mais il est clair qu'on ne leur attribuerait pas la possession normale d'une compétence sémantique ou de pensées.

Un déflationniste quant au holisme pourrait, parallèlement au déflationnisme quant à la vérité et à la signification (§ 5.3), soutenir qu'*il n'y a rien de plus*, dans les propriétés holistiques de la signification et des contenus mentaux, que ces platitudes. Il souscrirait donc au principe de compositionnalité, au holisme de la phrase, et au holisme méthodologique de l'attribution des significations et des croyances. Mais il refuserait de souscrire au holisme du langage et de la signification au sens de (ii) ci-dessus. Rien, dans les trois platitudes mentionnées, ne semble aller au-delà de cette forme minimale ou faible de holisme, et il n'y a rien en elles que semble devoir contester un critique du holisme comme Dummett, dans la mesure même où une théorie molécullariste de la signification repose sur de tels principes. Ce que contestent les critiques du holisme, c'est l'inférence, qu'ils jugent presque irrésistible (« sorite »), de ces platitudes à des thèses telles que : il y a une interdépendance *intrinsèque* entre les significations, les croyances et les comportements en sorte qu'on ne peut *jamais* désimpliquer leurs rôles respectifs ; le sens d'une expression dépend de celui de toutes les autres ; pour comprendre le sens de *a est F*,

1. Schiffer imagine un locuteur, Harvey, qui possède une théorie sémantique non compositionnelle pour son langage. Mais il n'attaque pas véritablement le principe de compositionnalité comme principe sémantique ; il entend montrer qu'une *théorie sémantique* vériconditionnelle n'est pas nécessairement compositionnelle, et appelle le principe « la Platitude ». Ni Dummett ni Fodor, les plus ardents critiques du holisme, ne le nient. Le principe de compositionnalité joue précisément un grand rôle dans la polémique de Fodor contre le connexionnisme (cf. Fodor et Pylyshyn, 1988).

il est nécessaire de comprendre le sens de *b est F*, de *c est F*, etc., mais aussi de *a est G*, de *b est G*, etc., *et ainsi de suite pour tout le langage et tous les contenus*. Il me semble cependant qu'il n'y a rien à contester dans le holisme si l'on se dispense d'inférer de telles conclusions des platitudes *(a)-(c)*, et qu'il y a une lecture possible de Davidson d'après laquelle il n'irait pas plus loin qu'un holisme minimal.

Mais j'ai admis aussi qu'il n'était pas évident que Davidson souscrive à un holisme seulement minimal de ce type. Nous devons donc esquisser une conception de la signification et des contenus mentaux qui à la fois préserve la vérité des platitudes du holisme minimal et interdise l'inflation menaçante du holisme radical. Je le ferai en m'appuyant sur la théorie des contenus sémantiques et psychologiques développée par Peacocke (1986, 1992), à laquelle j'ai déjà fait allusion à la fin du chapitre précédent.

La première étape, dans la formulation d'une telle conception, consiste à refuser une lecture radicale, que défend peut-être Davidson, de la platitude (*a*), affirmant que l'interdépendance des croyances et des significations est intrinsèque, c'est-à-dire qu'il n'est jamais possible d'isoler une signification indépendamment des croyances qu'entretient un individu, ni d'isoler une croyance indépendamment des significations de ses mots. Dire que nous ne pouvons pas interpréter une croyance d'un agent sans être capable au moins d'interpréter certaines de ses phrases est une chose. Soutenir qu'il n'y a pas de croyances ou de pensée sans langage en est une autre. La première affirmation est une vérité importante quant à l'interprétation des pensées, alors que la seconde est une thèse beaucoup plus contestable. Davidson (1975, 1982) semble passer de l'une à l'autre explicitement quand il soutient que seuls des êtres chez lesquels on peut discerner une certaine structure rationnelle et une certaine trame cohérente d'attitudes propositionnelles, et qui sont capables de communication linguistique, ont des croyances, et que les animaux n'en ont pas. Il semble soutenir que le langage seul et la communication rend possible la pensée, et qu'il y a en ce sens une interdépendance intrinsèque entre le langage et la pensée[1]. Son argument dépend ici de son principe transcendantal : signifier quelque chose, et avoir une croyance suppose qu'on soit radicalement interprétable. Mais tant que l'interprétation radicale est tenue comme s'appliquant nécessairement à des *phrases* tenues pour vraies et à des croyances qu'elles expriment, c'est faire une pétition de principe en faveur de la thèse selon laquelle la pensée dépend intrinsèquement du langage que de soutenir que seules les pensées susceptibles d'être articulées dans une structure *propositionnelle* et assertées linguistiquement peuvent être interprétées. Or il n'y a aucune raison de le supposer. La conception davidsonienne de la croyance fait de celle-ci une relation essentielle à des phrases (cf. § 1.5), mais il n'est pas évident que toute croyance soit une pensée susceptible d'être acceptée ou tenue pour vraie. Il n'entre pas ici dans mon propos de défendre cette thèse, mais il ne fait pas de doute que si nous voulons bloquer la source la plus radicale du holisme, nous devrons admettre qu'il n'est pas unilatéralement vrai qu'il y ait une interdépendance intrinsèque de la pensée par rapport au langage, et inversement (cf. Peacocke, 1986, chap. 8).

1. J'ai examiné cet argument dans Engel, 1992, chap. 4.

Si nous rejetons le principe de l'interdépendance des croyances et des significations en son sens holistique fort, nous devrons aussi, dans une seconde étape, élargir la base psychologique de l'interprétation. J'ai déclaré, au chapitre 3 *(in fine)* que cet élargissement n'est pas incompatible avec la théorie de l'interprétation radicale, qui pouvait inclure des hypothèses sur la psychologie des locuteurs plus substantielles que celles qui portent sur des croyances et des préférences. Mais Davidson ne nous dit pas comment cet élargissement est possible, et même dans ses écrits plus récents (1990), il s'en tient à sa stratégie minimaliste inspirée de la théorie de la décision seule, et cette stratégie est elle-même inspirée par son refus de considérer, dans l'interprétation des contenus mentaux, autre chose que les croyances, les désirs et leurs causes. C'est notamment la raison pour laquelle il rejette le recours à des contenus mentaux tels que des expériences, des sensations ou des perceptions (§ 6.1), parce que ces contenus lui semblent entraîner l'existence d'intermédiaires entre les phrases (et les croyances) et la réalité. Mais il n'est pas évident qu'on puisse interpréter le comportement d'un agent indépendamment des perceptions et des expériences qu'il peut posséder. Peacocke (1981, 1983, 1986), à la suite de Evans (1982), a soutenu qu'une théorie générale du contenu de pensées

devait intégrer les contenus perceptifs, et que l'explication de l'action notamment devait faire appel à des contenus « démonstratifs »[1]. Mc Ginn (1986) a soutenu qu'on devait concevoir une procédure d'interprétation à deux étapes, l'une assignant aux sujets des contenus d'expérience, l'autre des contenus de croyances et des significations. Dans quelle mesure une théorie de l'interprétation radicale peut-elle intégrer ces modifications sans perdre sa radicalité, c'est-à-dire sans perdre ce qui, aux yeux de Davidson, est sa raison d'être même ? Toute sa procédure repose sur l'idée que l'on doit pouvoir isoler des attitudes par rapport à des phrases *sans* individualiser au préalable des contenus de croyance ou d'autres contenus sémantiques, et sans délimiter une classe de contenus observationnels ou d'expérience indépendants des croyances et des attitudes propositionnelles, et de telles approches abandonnent ce réquisit. Mais d'un autre côté, comme le remarque Laurier (1994), c'est bien aussi son choix du tenir-pour-vrai (ou du préférer-vrai) comme concept empirique de base qui semble être à la source de son holisme linguistique et psychologique. Si nous voulons bloquer ce holisme, il nous faut donc renoncer à une base empirique aussi étroite. La question difficile est celle de savoir comment le faire sans souscrire à des formes d'atomisme et de molécularisme radicales comme celles que nous avons discutées.

Il est caractéristique de la théorie de l'interprétation de Davidson que, comme le dit Bilgrami (1992 : 183), elle ne soit pas sensible aux « conceptions » des agents, c'est-à-dire à ce que, dans un vocabulaire fregéen, on appellerait les « modes de présentation » des référents des prédicats et des noms d'un langage. Non qu'elle ignore la dimension fregéenne du sens, comme on l'a vu, mais parce qu'elle la contourne en formulant des théories de la vérité qui, dans les conditions appropriées, font seulement « office » de théories du sens, sans *être* des théories d'entités individualisées aussi finement que des sens fregéens. Mais il n'est pas évident que cette stratégie puisse résoudre les difficultés traditionnelles qui se posent en

1. Peacocke, 1983 : 78 sq., propose également des contraintes plus spécifiques que celles de Davidson sur l'interprétation des états psychologiques, comme la « contrainte d'étroitesse » (*tightness*) selon laquelle un ensemble d'attitudes ne doit pas être atribué à un sujet si un ensemble plus resserré, ou plus fin, peut aussi lui être attribué. Cela implique notamment que si l'on peut attribuer des croyances *et* des états perceptifs plus individualisants de la psychologie d'un agent, on devra choisir d'attribuer les seconds. Cf. aussi Campbell 1986.

philosophie du langage notamment à propos de la référence des noms propres[1]. Une théorie des contenus qui les individualiserait plus finement que celle de Davidson devra donc sans doute réintroduire quelque chose comme des « modes de présentation » fregéens. C'est du moins ce que soutient Peacocke (1983, 1986).

Le cadre d'analyse proposé par Peacocke (1992) satisfait aux conditions qui viennent d'être énoncées. Peacocke entend fournir ce qu'il appelle une théorie des concepts, entendus comme composant de pensées, celles-ci étant elles-mêmes entendues au sens « fregéen » d'entités structurées, susceptibles d'être vraies ou fausses. Selon Peacocke, la nature d'un concept est donnée par une analyse de ses « conditions de possession », c'est-à-dire des conditions nécessaires et suffisantes pour la saisie de ce concept. (En d'autres termes, les conditions de possession sont, au niveau des concepts, l'équivalent des « conditions d'acceptation rationnelle » pour les pensées complètes.) Peacocke tient ce réquisit pour équivalent à celui de manifestabilité de la compréhension de Dummett. La forme générale d'une analyse des conditions de possession pour un concept doit satisfaire à ce que Peacocke appelle la condition A(*C*) :

> A(*C*) : Un concept *F* est le concept unique *C* que possède un sujet qui doit satisfaire à la condition A(*C*).

Les conditions de possession qui individualisent les concepts peuvent dans certains cas (ceux des concepts logiques) requérir que les sujets trouvent

1. Je pense ici particulièrement au « problème de Kripke », analysé par Bilgrami, 1992, *passim*. Bilgrami, dans ce livre (chap. 4), et dans 1992b, propose une analyse du problème originale, qui le conduit à soutenir que le holisme est une doctrine inoffensive. Bilgrami distingue deux niveaux : celui d'une théorie de la vérité, qu'il appelle « agrégatif », et celui de son application à l'explication du comportement (linguistique ou non), qu'il appelle « local ». Au premier niveau, agrégatif, le holisme règne : un interprète, quand il formule une théorie-T pour le langage d'un locuteur, ne peut manquer d'envisager *toutes* les croyances et significations possibles. Mais au second niveau, local, l'explication du comportement doit restreindre l'ensemble initial pour rendre compte des « conceptions » de l'agent. Il s'ensuit que le holisme est trivial, et inoffensif, car les produits de l'interprétation n'interviennent qu'au second niveau. L'erreur d'auteurs comme Fodor est de négliger cette distinction de niveaux. En substance, cette distinction recoupe celle de Davidson entre ce qui est interne à la théorie et l'explication de la théorie, invoquée ci-dessus pour répondre au problème holisme. Mais tant que Bilgrami ne nous dit pas exactement quelles contraintes pèsent sur l'explication du comportement, ni comment on doit réviser la théorie de l'interprétation radicale pour rendre compte du niveau « local », il ne lève pas les doutes antiholistes. J'examine plus en détail cette proposition dans Engel (à paraître). Sur le holisme de Davidson, cf. aussi Heal 1986.

certaines inférences impliquant ces concepts « primitivement irrésistibles ». Ainsi, pour un concept *C*, si un sujet trouve évidentes les transitions de *p*, *q*, à *p C q*, de *p C q* à *p*, et de *p C q* à *q*, le concept en question est celui de conjonction[1]. Dans d'autres cas, celui des expériences perceptives par exemple, les conditions de possession différeront. Peacocke requiert également que les concepts satisfassent à une forme de principe compositionnalité pour les pensées, c'est-à-dire que la valeur sémantique ou la référence d'un concept est déterminée par la valeur sémantique ou la référence de ses parties. Peacocke admet une forme de molécularisme, comparable à celui de Dummett, pour les concepts : il suppose que les conditions de possession d'un type donné de concept ne peuvent en général présupposer les conditions de possession d'un autre type de concept si ces dernières présupposent les conditions de possession du concept de premier type et donc qu'il existe une hiérarchie dans les concepts (1992 : 12). On bloque ainsi une forme de holisme, parce qu'il n'est pas vrai que les conditions de possession d'un concept déterminé doivent présupposer celles de *tous* les autres concepts, sans restriction. Mais il admet aussi qu'il existe des « holismes locaux », par lesquels certains concepts se trouvent dépendants d'autres (*ibid.*, 10). Peacocke admet aussi que le holisme est « correct au niveau des concepts » (*ibid.*, 52) au sens suivant :

> (CG) Si un sujet peut entretenir la pensée *Fa* et posséder également le mode de présentation *b* qui fait référence à quelque chose dans le domaine des objets dont le concept *F* est vrai ou faux, alors le sujet a la capacité conceptuelle de former des attitudes propositionnelles contenant le contenu *Fb*. (*ibid.* : 42)

qui n'est autre qu'une version de ce que Evans (1982 : sec. 4.4) appelle la *Contrainte de généralité*, et qui correspond à notre platitude (c) concernant la généralité et la productivité ci-dessus.

Ces conditions valent au niveau des concepts comme constituants de pensées. Comme telles, elles présupposent une relative indépendance de la pensée par rapport au langage, et par conséquent impliquent un rejet de l'interdépendance des pensées et des significations au sens radical.

1. Pour un exposé des analyses de Peacocke concernant les constantes logiques, cf. Engel, 1989, chap. XII (édition anglaise).

Comment s'appliquent-elles à présent aux significations ? Il n'est pas vrai, nous dit Peacocke, qu'à tout concept doit correspondre un mot dans un langage (*ibid.*, 3). Mais on doit reconnaître que la maîtrise d'un mot peut être une façon d'acquérir un concept. Il peut y avoir des *conditions d'attribution* d'un concept qui fassent référence à l'assentiment que donne un sujet à des *phrases* contenant un mot, qui soient distinctes des conditions de possession de ce concept, mais se trouvent *associées* à des conditions de possession (*ibid.*, 29 sq.). Il n'est donc pas vrai en général que la possession d'un concept suppose la maîtrise des mots et des phrases d'un langage exprimant ce concept et ses modes de composition. Mais si l'usage d'un mot est tel que le locuteur répond à certaines conditions de possession du concept que ce mot exprime, les réquisits qui valent pour les concepts dans la pensée seront valables aussi pour les mots et leurs significations dans le langage. Peacocke soutient qu'en ce sens il échappe à l'objection de Dummett (cf. § 4.5) selon laquelle une théorie qui déterminerait les conditions de possession d'un concept au contenu d'une pensée *indépendamment* du langage souscrirait nécessairement à une conception psychologiste du langage comme *code* pour les pensées (*ibid.*, 34). Il rejette aussi la position de McDowell (cf. § 5.4) selon laquelle une théorie de la signification ne peut pas être autre que modeste, et doit toujours *présupposer*, dans l'individuation des pensées et des significations, ces pensées et significations mêmes. Il ne s'agit ni de dire que la pensée est strictement antérieure au langage, ni qu'elle en dépend strictement.

Une théorie du contenu de ce type est hautement sujette à controverse, et bien des questions devraient recevoir une réponse avant de considérer qu'elle puisse être établie. Mais ce n'est pas ce qui m'occupe ici. J'ai seulement voulu soutenir que l'on pouvait formuler une théorie qui satisfasse les trois platitudes du holisme minimal — le holisme méthodologique des attributions de croyance et de signification, le holisme de la phrase et le principe de compositionalité, et la condition de généralité et de productivité — *sans* souscrire au holisme radical de la signification et des croyances, et *sans* adopter une forme radicale d'atomisme ou de molécularisme. On admet donc le holisme minimal, mais on ne s'en tient pas à lui : on admet un holisme limité propre à la théorie des conditions de possession des concepts. L'argument est ici exactement parallèle à celui

qui nous avait fait passer, au chapitre précédent, du déflationnisme ou du quiétisme quant à la vérité et la signification, au réalisme minimal. Si nous nous rappelons que la théorie de Peacocke satisfait aussi aux conditions du réalisme minimal, alors nous pourrons considérer qu'elle satisfait aussi aux conditions épistémiques sur la signification, c'est-à-dire à l'exigence qu'une théorie de la signification soit aussi une théorie de la compréhension. C'est la dernière tâche à laquelle nous devons à présent nous atteler.

7

Comprendre un langage

> Do you speak Spanish ? — I don't know. I've never tried.
>
> P. G. Wodhouse.

7.1. Théorie de la signification et théorie de la compréhension

L'essence de la critique antiréaliste de la conception véri-conditionnelle de la signification est que cette conception ne peut pas être une véritable théorie de la compréhension du langage, puisqu'elle associe cette compréhension à des conditions de vérité inconnaissables. J'estime avoir répondu, à la fois au nom de la conception davidsonienne et au nom de ce que j'ai appelé réalisme minimal, à la partie métaphysique de cette critique, en soutenant que le réalisme des conditions de vérité n'entraînait pas le réalisme externe qu'attaque principalement Dummett. J'ai également soutenu que le holisme de Davidson satisfaisait largement aux conditions de manifestabilité du sens énoncées par les théoriciens antiréalistes, et qu'il y a une version du holisme, distincte de celle de Davidson, parfaitement acceptable. Mais la partie n'est pas pour autant gagnée pour le défenseur d'une théorie réaliste de la signification comme celle que j'entends proposer. Car le fait qu'une théorie réaliste de la signification n'ait pas les implications métaphysiques qu'elle est supposée avoir ne montre pas que les doutes initiaux de l'antiréaliste quant à sa capacité à articuler une théorie de la compréhension du langage ne sont pas fondés. Le problème d'une épistémologie de la théorie de la signification reste entier : quelle sorte de théorie de la compréhension du langage la conception véri-conditionnelle a-t-elle à nous offrir ?

J'ai admis, en formulant le réalisme minimal au chapitre 5, que cette conception devait incorporer des contraintes épistémiques sur la signification. Cela revient à accepter une forme générale du réquisit de manifestabilité de Dummett : *une théorie de la signification doit être une théorie de la compréhension*. En son sens général (qui ne présuppose pas les restrictions très strictes de Dummett), ce principe veut simplement dire qu'une théorie de la signification ne doit pas produire un résultat si éloigné des propriétés de la cognition humaine et de la compétence linguistique usuelle des locuteurs qu'on puisse se demander en quoi ceux-ci peuvent bien connaître cette théorie. C'est presque une platitude. Mais notre lecteur est à présent familiarisé avec la nécessité d'interpréter les platitudes. Il y a trois manières de lire ce principe très général[1]. Comme toute équivalence, elle peut être lue de trois manières. En premier lieu, on peut la lire de gauche à droite, en donnant la priorité à la théorie de la signification, c'est-à-dire en supposant que les faits énoncés par une théorie de la signification établiront les faits d'une théorie de la compréhension. C'est la lecture davidsonienne. Une théorie de la signification qui fournit des interprétations des énoncés d'un locuteur, en assignant des conditions de vérité aux phrases de son langage, énoncera ce qui est connu par quiconque est capable de comprendre le langage. Mais elle laissera de côté les manières dont cette compétence est acquise et maîtrisée. Appelons, à la suite de B. C. Smith (1992), cette conception le point de vue *interprétatif*. En second lieu, on peut soutenir que ni l'un ni l'autre des termes de l'équivalence n'a de priorité : les faits relatifs à la signification dans le langage et les faits relatifs à la compréhension doivent être établis simultanément par une théorie qui rend justice aux deux types de faits. Une théorie de la signification devra à la fois être sensible à ce qui est connu et à la manière dont ce qui est connu est connu, c'est-à-dire à la manière dont les locuteurs manifestent ce savoir dans l'usage. Cela correspond aux réquisits mis en avant par l'antiréalisme dummettien. Appelons cette conception le point de vue *descriptif*. Enfin, et en troisième lieu, on peut lire l'équivalence de droite à gauche, et soutenir que les faits relatifs à la compétence sémantique des locuteurs imposent des contraintes

1. Comme l'a bien vu B. C. Smith (1992), dont je suis ici la remarquable présentation.

particulières sur le choix d'une théorie de la signification, ou tout au moins que ce choix doit refléter ces faits. Ceux-ci seront essentiellement de nature psychologique. Une théorie de la signification devra donc être fidèle aux propriétés psychologiques et cognitives de la compétence sémantique. On peut considérer que l'approche de Chomsky et de ses disciples est conforme à ce point de vue, que nous pouvons appeler *explicatif*. Chacun de ces points de vue est légitime, mais ils entrent potentiellement en conflit les uns avec les autres. Le point de vue interprétatif rend compte de l'aspect intersubjectif et public du langage, et du caractère des attributions à la troisième personne de significations. Mais il risque du même coup de perdre l'aspect de connaissance des significations à la première personne, et les faits relatifs à la cognition du langage sur lesquels insistent le second et le troisième point de vue respectivement. Le point de vue descriptif rend compte du second aspect, mais tend à négliger le premier et le troisième. Et le point de vue explicatif rend justice au troisième sans nécessairement prêter attention aux deux premiers. Dans une large mesure, un certain nombre de débats contemporains en philosophie du langage proviennent du fait que l'on se place au départ de l'un ou l'autre de ces points de vue, et un grand nombre des tensions, des difficultés et des apories que nous avons analysées dans ce livre proviennent de l'adoption implicite d'un des points de vue au détriment des autres. Et pourtant chacun semble avoir sa légitimité, et on peut rêver d'une conception qui rende justice à chacun d'eux. Mais nous avons déjà vu que ce n'est pas une tâche facile. Le débat est encore compliqué par la possibilité d'une autre perspective. Chacun des points de vue qui viennent d'être mentionnés présuppose ce que nous avons appelé, à la suite de Wright, *l'objectivité de la signification*. On suppose que si théorie de la signification il y a, elle doit porter sur quelque chose d'objectif, de réel, et de déterminable. Mais on peut envisager une quatrième espèce de théoricien, qui vienne nous dire : quelle garantie avez-vous qu'il existe quelque chose comme des *faits* de signification dont il puisse y avoir une « théorie » ? comment savez-vous qu'il existe des propriétés telles que *signifier quelque chose par une expression* ? en ce sens comment pouvez-vous souscrire à un *réalisme* quant à cette notion ? La menace introduite par ce genre de question existe au niveau de chacun des points de vue envisagés. Elle existe pour

la position interprétative sous la forme de l'indétermination de la traduction ou de l'interprétation. Elle existe, du point de vue descriptif, sous la forme du défi que l'antiréaliste pose au réaliste, quand il le somme d'exhiber les faits de compréhension correspondant à la saisie de phrases transcendantes ou face au holisme de la signification. Enfin elle existe pour le point de vue explicatif, parce qu'il semble qu'un style d'explication « galiléen » qui expliquerait la compétence sémantique par des faits naturels de la cognition humaine devra laisser de côté l'aspect *normatif* de l'usage du langage. Mais des normes peuvent-elles être expliquées par des faits, ou réduites à eux ? Cette menace, que nous pouvons appeler *sceptique*, surgit de ce que l'on a appelé les « considérations wittgensteiniennes sur suivre-une-règle », que Kripke (1981) a mises en avant.

Je voudrais ici poser la question de savoir quelle forme pourrait prendre une théorie satisfaisante de la compréhension qui puisse intégrer ces différents points de vue et répondre à ces doutes, et essayer de proposer une telle conception. Mais il serait bien trop ambitieux d'envisager tous les aspects de ce que l'on peut appeler la connaissance du langage. Je considérerai principalement l'idée selon laquelle la connaissance révélée par une théorie de la signification est une connaissance « tacite » ou « implicite ». Aucun des théoriciens de la signification que nous avons considérés ne tient la compréhension du langage révélée par une théorie de la signification comme une connaissance *explicite*, propositionnelle, au sens usuel. Les axiomes et les théorèmes d'une théorie-T ne peuvent pas être supposés faire partie de l'équipement conscient normal d'un locuteur. La difficulté avait été bien formulée par Strawson (1976) au sujet de l'analyse davidsonienne de la forme logique des phrases d'action : s'il est vrai que la forme logique réelle de ces phrases est celle d'une quantification sur des événements, alors même que la manière usuelle, explicite, dont nous comprenons la structure de surface de ces phrases ne coïncide pas avec ce que prescrit cette forme logique, en quoi la proposition de Davidson peut-elle nous révéler *ce que nous comprenons* quand nous comprenons ces phrases ? Le davidsonien est contraint de dire ici que ce n'est pas une connaissance explicite de la forme logique que nous avons. De même, comme on l'a vu, Dummett considère que la connaissance du langage est dans une large mesure « implicite ». Et la perspective chomskyenne

suppose que la majeure partie de la connaissance du langage est « tacite ».[1] Cette idée est commune aux trois perspectives envisagées. Mais on va voir que ce n'est pas dans le même sens dans chaque cas, et que la plupart des difficultés qui viennent d'être évoquées se concentrent sur l'emploi de cette notion.

7.2. Le point de vue interprétatif et la connaissance tacite

Si Davidson assigne bien comme tâche à une théorie de la signification d'élucider la compréhension qu'un locuteur a d'une langue naturelle, il ne soutient pas qu'une théorie de ce genre représente la compétence sémantique *effective* d'un locuteur, au sens où cette compétence serait constituée, on peut le présumer, par des états psychologiques réels. Une théorie-T, adjointe à une théorie de l'interprétation, n'est pas destinée à nous fournir une analyse des états, psychologiques, neurophysiologiques ou cognitifs, qui peuvent constituer le *processus* de la compréhension des phrases d'une langue. Il ne s'agit pas de nier qu'il y ait de tels états ou processus, ni qu'ils jouent un rôle important dans la compréhension du langage. Davidson soutient, en vertu de son monisme anomal, qu'il y a de tels états psychologiques, et qu'ils sont identiques à des événements physiques du cerveau, mais il rejette l'idée qu'on puisse établir une corrélation *nomologique* entre les contenus sémantiques qu'établirait une théorie de la signification, et des états psychologiques, cognitifs ou neurophysiologiques. Que peut bien alors représenter une telle théorie ? La réponse la plus naturelle est qu'elle ne représente pas la forme psychologique de la compréhension, mais la forme ou la structure sémantique du langage, indépendamment des propriétés réelles de ceux qui le parlent, c'est-à-dire une certaine structure abstraite.

Selon cette interprétation, la nature d'une théorie sémantique n'est pas de caractériser les traits de la psychologie réelle des locuteurs, mais seulement la forme abstraite des relations sémantiques à l'œuvre dans le

1. Dans la littérature récente, c'est évidemment Chomsky et ses disciples qui ont articulé cette notion le plus précisément. Mais c'est aussi une vieille idée. Voir en particulier les articles sur le *Ménon* recueillis dans M. Canto (1991), et notamment l'article de Davidson.

langage lui-même, considéré comme un objet abstrait. Dans ces conditions, toute théorie qui pourrait représenter correctement la structure sémantique d'un langage serait *par là même* adéquate, et si la forme de ces théories doit être vériconditionnelle, il n'y a, de ce point de vue, aucune raison *a priori* de privilégier telle théorie plutôt que telle autre. Une sémantique fondée sur la théorie des modèles, par exemple, pourrait répondre à cet objectif tout autant qu'une sémantique de type tarskien. Appelons cette conception celle de l'*autonomie* totale de la sémantique par rapport à une théorie de la compréhension du langage. C'est par exemple la conception de Montague (1974) et celle de Cresswell (1976).

Ce n'est cependant pas, comme on l'a vu (§ 1.1), l'interprétation correcte du projet de Davidson. Selon lui, au contraire, les propriétés formelles d'une théorie sémantique doivent précisément être contraintes par la manière dont elles peuvent rendre compte de l'apprentissage d'un langage, de sa compréhension, et de son interprétation. Le choix du format tarskien, la condition de finitude des axiomes, et les diverses conditions de la théorie de l'interprétation *sont* destinées à rendre compte de ces propriétés. Mais de quelles propriétés peut-il s'agir, si ce ne sont pas non plus les propriétés réelles des locuteurs ? Ce ne peuvent être que des propriétés *idéales* de la compréhension : une théorie sémantique est destinée à caractériser *ce que* les locuteurs *devraient* connaître s'ils comprennent leur langage, et une théorie de l'interprétation établit également *ce qu'*ils ont à connaître pour interpréter. Mais ces théories n'ont rien à nous dire sur la question de savoir *comment*, en fait, les locuteurs comprennent et interprètent. Comme on l'a vu au chapitre 3, Davidson considère que les aptitudes manifestées par les interprètes dans l'usage effectif de la communication relèvent de stratégies, de techniques et d'aptitudes qui ne se laissent pas systématiser. Il nie même qu'il y ait, dans la communication, un objet commun, le langage, qu'une interprétation puisse représenter. Il s'ensuit qu'une théorie de la signification n'a pas à être une théorie de la compréhension psychologique effective des locuteurs, et qu'il n'y a aucune raison de supposer que ceux-ci connaissent, au titre de leur équipement cognitif ou psychologique, une théorie-T ou un langage lui correspondant. C'est le théoricien, l'interprète, qui *caractérise* une telle compétence : celle-ci n'est pas une propriété du locuteur lui-même. Dans

ces conditions, cela n'a pas plus de sens de dire qu'une théorie de la signification pourrait être connue par un locuteur au sens où celui-ci pourrait articuler explicitement les propriétés qu'établit une théorie de la signification, que de dire qu'il connaît ces propriétés « implicitement ». C'est ce raisonnement que fait Foster quand il présente le projet davidsonien :

> Plutôt que de demander un énoncé de la connaissance implicite dans la compétence linguistique, demandons un énoncé d'une théorie dont la connaissance suffirait pour une telle compétence. Au lieu de demander un énoncé des faits métalinguistiques que la maîtrise du langage reconnaît implicitement, demandons un énoncé des faits dont la reconnaissance nous fournit la maîtrise. Ce que nous demandons alors est toujours une théorie de la signification, mais déchargée de l'hypothèse problématique selon laquelle celui qui a maîtrisé le langage a, à un niveau profond quelconque, absorbé l'information qu'elle fournit. La théorie révèle la machinerie sémantique mise en jeu dans la compétence, mais laisse indéterminée la forme psychologique dans laquelle se trouve réalisée la compétence elle-même (*ibid.*, 2).

Cela semble bien correspondre à la position davidsonienne officielle :

> Une théorie de la vérité décrit les conditions sous lesquelles un énoncé fait par un locuteur est vrai, et ainsi ne dit rien directement quant à ce que sait le locuteur... Une théorie de la vérité lie le locuteur à l'interprète : elle décrit d'emblée le comportement linguistique réel et potentiel du locuteur et donne la substance de ce qu'un interprète compétent sait quand il peut saisir les significations des énoncés du locuteur (1990 : 311-312, cf. aussi 1986 : 438).

Mais cette position officielle conduit à des conséquences inacceptables si l'on admet le réquisit de manifestabilité au moins sous sa forme faible énoncée plus haut : selon cette position, une théorie de la signification est une théorie de ce que sait *l'interprète*, pas de ce que sait le locuteur, et elle réduit la compréhension du langage à la compréhension qu'a un tiers de ce qu'un locuteur comprend. C'est précisément, comme on l'a vu, la substance de la critique de Dummett (§ 4.5) : selon lui, la conception davidsonienne a pour effet de divorcer totalement signification et compréhension[1].

1. C'est le point de vue de Smith (1992) qui va jusqu'à qualifier la position de Davison d'« éliminativiste » et d'« épiphénoméniste » quant au langage. Il n'y aurait, selon cette lecture, rien dont la théorie de la signification serait la théorie. En un sens, c'est aussi la lecture de Ramberg (1987). Mais Smith a tort, à mon sens, de négliger les arguments que j'avance ci-dessous, et de prendre trop au pied de la lettre les critiques de Dummett.

Mais la position officielle est-elle bien la position de Davidson ? Si elle l'est, elle se rapproche dangereusement de la thèse de l'autonomie totale de la sémantique. Or Davidson laisse entendre, dans le passage cité ci-dessus, que si la théorie ne dit rien « directement » de ce que sait le locuteur, elle peut en dire quelque chose *indirectement*. A nouveau cela peut être compris comme l'expression du fait que les propriétés de la compréhension effective d'un locuteur sont déférées à un interprète et à une communauté d'interprètes, et donc parfaitement sous-déterminées relativement au locuteur lui-même. Cela ne nous permettrait pas plus d'assurer les bases psychologiques de la compréhension. Il y a bien une lecture de Davidson conforme à cette interprétation, et ouverte à cette critique. Mais il y a une autre lecture possible, à mon sens la bonne, d'après laquelle une théorie de la signification pourrait caractériser *indirectement* la compétence psychologique réelle d'un locuteur. Après tout, Davidson lui-même ne s'interdit pas de dire sur quelles *sortes* d'états psychologiques repose notre connaissance des significations : ce sont des intentions, des croyances, des désirs, et d'autres attitudes propositionnelles, qui sont des états psychologiques réels, doués de contenu, et non réductibles à des dispositions comportementales ou à des états physiques. Si ces états sont ce que l'interprète assigne au locuteur, ils jouent un rôle non trivial dans ce que l'on doit considérer comme le processus de la compréhension. Le problème se pose alors de savoir comment ils se rapportent aux structures abstraites d'une théorie sémantique. Il y aussi une autre raison de résister à l'interprétation officielle. Elle provient de l'idée, analysée au chapitre 1, de structure sémantique. Si une théorie sémantique avait seulement pour but de décrire *l'information* contenue dans une théorie-T, que doit maîtriser un locuteur idéal, sans chercher à dire en quoi cette théorie peut être apprise et comprise, quelle place resterait-il pour l'idée que la théorie doit articuler la structure, compositionnelle et récursive, d'une langue ? Une théorie de l'espèce de T1 ou T2 du § 1.3 décrirait bien ce qu'un locuteur *devrait* connaître. Mais elle serait, comme on l'a vu, inadéquate pour rendre compte du caractère structuré de la compétence sémantique. Le réquisit de finitude des axiomes lui-même n'a pas un sens psychologique, mais il impose des conditions précises à toute théorie psychologique de la compétence (Davidson, 1965).

Supposons qu'il soit possible d'apprendre le jeu d'échecs à des enfants avant qu'ils soient capables de parler et d'articuler verbalement les règles de ce jeu, et qu'ils jouent correctement. Il sera alors naturel de leur attribuer une *connaissance* des règles. Cette connaissance, par définition, n'est ni propositionnelle ni explicite, et il semble légitime de dire qu'elle est « tacite » ou « implicite ». Mais, dans ce cas, attribuer une connaissance implicite des règles ou de la « théorie » des échecs n'est rien d'autre qu'une manière oblique de décrire leur comportement. Cela laisse totalement indéterminée la nature des états qui constitueraient leur compétence. De même avec une théorie de la signification, à cette différence près que l'on a affaire à des sujets au moins en partie capables d'articuler certains des principes de leur compétence sémantique. La notion de connaissance tacite demeure donc triviale tant que l'on ne spécifie pas la manière dont l'information contenue par une théorie sémantique pourrait être « réalisée » dans la psychologie ou dans le comportement effectif des locuteurs, de manière à révéler des propriétés jouant un rôle *causal* quelconque dans la compétence[1]. Elle serait non seulement triviale, mais absurde, pour au moins deux raisons. Disons-nous, tout d'abord, que du fait que nous avons axiomatisé l'arithmétique élémentaire au moyen des axiomes de Peano, les axiomes *décrivent* notre compétence arithmétique usuelle ? C'est commettre l'erreur de confondre ce que systématise une théorie avec les processus psychologiques effectifs impliqués dans sa compréhension. En second lieu, en vertu de la méthodologie de l'interprétation, plusieurs théories-T extensionnellement équivalentes pourraient être attribuées à un locuteur. Faudrait-il dire qu'il connaît tacitement ces diverses théories ?

Doit-on alors renoncer à dire qu'une théorie sémantique pourrait jouer un rôle explicatif quelconque pour le mécanisme de la compréhension ? Non, car, comme on l'a déjà dit, la contribution d'une théorie sémantique doit avoir un rapport avec la manière dont elle représente la structure sémantique d'un langage. Mais nous devons dire de quelle sorte de structure sémantique nous parlons. On peut, à la suite de Davies (1981, 1986), distinguer quatre notions de structure sémantique. Il y a, aux deux extrêmes

1. Cf. Wright, 1987 : 220-221, cf. aussi Wright, 1981 : 110.

de la classification, une structure *abstraite*, caractérisée ci-dessus par la thèse de l'autonomie, et une structure *psychologique*. La première est purement formelle et il y a en ce sens autant de structures différentes possibles pour un langage. La seconde concerne la psychologie d'un locuteur particulier, avec ses caractéristiques éventuellement idiosyncrasiques. Mais il y a, entre ces deux extrêmes, deux autres notions de structure. L'une est abstraite, mais en un sens distinct de la première : elle caractérise la compétence sémantique d'un locuteur idéal, et correspond, selon Davies, à la condition suivante :

> Si quelqu'un pouvait (par des moyens rationnels inductifs) venir à savoir ce qu'une phrase *s* signifie purement sur la base de sa connaissance de ce que d'autres phrases $s_1 \ldots s_n$ signifient, alors une théorie sémantique devrait révéler la signification de *s* comme le produit des mêmes ressources que celles qui sont déployées dans $s_1 \ldots s_n$. (Davies, 1986 : 132).

L'autre notion possible est psychologique, mais ne s'identifie pas pour autant à la structure psychologique individuelle d'un locuteur. Elle concerne la compétence d'un locuteur normal du langage, indépendamment des propriétés des locuteurs individuels. Selon cette conception, la structure axiomatique d'une théorie-T et la structure de dérivation de ses théorèmes *reflètent* la structure causale réelle qui sous-tend la compétence psychologique normale. Elle correspond à ce que Davies (1981 : 53 ; 1986 : 133) appelle la « condition du reflet » (*Mirror Constraint*) : elle stipule que les structures cognitives effectives du locuteur normal reflètent les structures dérivationnelles spécifiées au sens ci-dessus (« idéal »). L'idée qui sous-tend cette condition du reflet est que l'investigation que nous pouvons faire sur la structure causale et psychologique de la compétence des locuteurs doit pouvoir être reflétée par la structure elle-même des axiomes, des dérivations, et des théorèmes de la théorie-T. L'isomorphie — en un sens encore à définir — des deux structures n'implique pas leur *identité* : on ne commet pas l'erreur de confondre structure sémantique et structure psychologique, et de croire que la contribution du sémanticien sera une contribution directe à la psychologie de la compréhension.

Quelle est, de ces quatre types de structure, celle qui correspond au projet davidsonien officiel ? Il est clair que c'est la seconde, celle qui

reconstruit rationnellement, comme le dit Wright (1987 : 216), la compétence sémantique idéale. Prise au pied de la lettre, elle est incompatible avec une théorie qui donnerait droit au quatrième type de structure et à la condition du reflet, et, en ce sens, la perspective interprétative doit être nécessairement divorcée de la perspective explicative. Mais il n'est pas interdit de chercher à explorer, à partir de celle-là, cette dernière option.

7.3. Evans sur la connaissance tacite

La condition du reflet pose deux types de problème. Le premier concerne la nature des états psychologiques « reflétés » par la structure sémantique. C'est une question empirique, du ressort du psycholinguiste ou du neuropsychologue. Le second problème n'est pas empirique, et porte sur la nature de la connexion qu'on peut établir entre les deux types de structure, et sur son unicité. C'est celui que nous avons déjà mentionné ci-dessus : comment peut-il y avoir des données empiriques qui puissent garantir l'attribution à un sujet d'une théorie sémantique plutôt que d'une autre, extensionnellement équivalente ? C'est le défi que Quine (1972) pose au grammairien chomskyen, quand celui-ci entend parler d'une connaissance tacite de la grammaire. Quine propose son utile distinction entre un comportement *conforme à une règle* et un comportement *guidé par la règle* : un comportement est conforme à une règle s'il peut être décrit correctement en mentionnant celle-ci ; un comportement est guidé par une règle si le sujet connaît la règle et peut l'observer. Selon Quine, le sens dans lequel Chomsky et ses disciples parlent de connaissance tacite est intermédiaire entre ces deux sens, puisque le comportement grammatical est supposé guidé, bien qu'inconsciemment et implicitement, par les règles. « Guider », dit Quine, n'est pas une simple question de conformité, mais une question de cause et d'effet. La « nouvelle tâche » du grammairien est de dire, de deux systèmes extensionnellement équivalents mais distincts de règles, lequel *guide* réellement la compétence. Et il suggère que si ce sens doit se rattacher à des dispositions observables du sujet, le chomskyen n'a pas montré comment effectuer ce choix.

Dans « Semantic Theory and Tacit Knowledge » (1981), Gareth Evans entreprend de formuler une notion de connaissance tacite qui réponde à ce défi quinien. Pour rendre explicite la différence entre deux théories distinctes mais extensionnellement équivalentes, il utilise l'exemple, déjà discuté au § 1.3, des deux théories T2 et T3 pour un langage L2 de 100 phrases composées toutes à partir de 10 noms et de 10 prédicats. (Pour simplifier, appelons cette fois-ci respectivement T2 « T1 », T3 « T2 » et L2 « L ».) L'une des théories, T1, a 100 axiomes-T spécifiant le sens de chaque phrase, alors que l'autre, T2, a 21 axiomes, un pour chaque nom et un pour chaque prédicat, plus l'axiome « compositionnel » indiquant le mode de composition d'une phrase à partir d'un nom et d'un prédicat. Le défi de Quine peut être formulé comme celui de choisir, de T1 et T2, laquelle attribuer à un sujet une connaissance des significations des phrases de L, et en ce sens il n'y a pas de raison de dire que le sujet connaît tacitement l'une plutôt que l'autre. Pourtant nous objecterons que T2 seule peut révéler la *structure* de la compréhension du sujet. Mais eu égard aux dispositions observables, ce choix semble arbitraire. Evans fait la proposition suivante :

> Je suggère que nous construisions la thèse que quelqu'un connaît tacitement une théorie de la signification comme attribuant à cette personne un ensemble de dispositions — une correspondant à chacune des expressions pour lesquelles la théorie fournit un axiome distinct... L'attribution de connaissance tacite... implique l'idée qu'il y a un état unique du sujet qui figure dans une explication causale de la manière dont il réagit avec cette régularité à toutes les phrases contenant l'expression (Evans, 1981 : 124-125, cité d'après 1985 : 328-329).

Evans ajoute que les dispositions doivent être entendues au sens « plein », comme états réels, pas au sens de pures régularités de comportement. Mais en quoi cette proposition nous permet-elle de justifier l'attribution de T2 plutôt que de T1 ? Après tout, T1 permet d'attribuer 100 dispositions distinctes, alors que T2 permet d'en attribuer 20. Et il est plus difficile, si l'on choisit T2, d'attribuer des dispositions isolées, car elles sont « interconnectées », et ne peuvent se manifester isolément. Mais il y aura, selon Evans, une différence empirique testable entre T1 et T2. Attribuer à un sujet qui entendrait par exemple la phrase « Michel Foucault est chauve », selon le modèle de T1, une connaissance tacite de la signification du nom

« Michel Foucault » et du prédicat « chauve », impliquera qu'il y ait *plusieurs* liens possibles *indépendants* entre la reconnaissance par le sujet de la phrase « *Fa* » et l'état de sa connaissance tacite représenté par T1, puisque le sujet connaîtra à chaque fois un axiome distinct. Au contraire, une attribution de connaissance tacite selon le second modèle (T2) n'impliquera qu'un seul lien de ce genre, en vertu de l'axiome compositionnel commun à toutes les phrases de la forme *Fx* (1985 : 330-331). L'attribution de connaissance tacite de T2 implique qu'il y ait vingt dispositions de ce genre. Cette différence, en elle-même purement théorique, entre T1 et T2, doit pouvoir être traduite empiriquement, par exemple en considérant la perte, momentanée ou définitive de la compétence relative à certaines phrases, des lésions cérébrales, ou les manières dont s'effectue la perception de telle expression. En ce sens, selon Evans, le défi de Quine peut trouver une réponse[1].

Mais cela ne répond (et seulement partiellement, on le verra) qu'à l'une des questions que l'on peut adresser à la notion de connaissance tacite, celle du lien possible entre structure sémantique et structure causale. Cela ne répond pas au doute principal : comment la « connaissance tacite » peut-elle être une *connaissance* ? Au sens usuel, connaître, c'est — en principe — savoir que l'on sait, et être capable d'articuler cette connaissance. Si un état de connaissance est une attitude propositionnelle, une relation d'un sujet à des contenus vrais, quel sens cela peut-il avoir de dire que l'on « connaît » tacitement des axiomes et des théorèmes d'une théorie-T ? Il y a bien un sens dans lequel la théorie de Evans selon laquelle les états de connaissance tacite correspondent à des dispositions pourrait être compatible avec l'ascription d'attitudes propositionnelles : c'est le sens où l'on dit que les croyances et autres attitudes sont elles-mêmes dispositionnelles, et distinctes d'états mentaux « occurrents ». Cette théorie se heurte à des difficultés classiques, et ce n'est pas celle de Evans. Mais sa théorie se heurte également au fait suivant : bien que chaque axiome corresponde à une disposition, on ne peut pas dire que le *contenu*

1. Cette réponse de Evans à Quine est largement parallèle à celle qu'il donne à l'argument de l'indétermination de la traduction dans « Identity and Predication », (1975 ; *in* Evans, 1985). Le lecteur français trouvera une analyse de cet argument dans Clementz, 1985. Cf. également mes analyses de la position de Evans dans Engel, 1985a.

sémantique de chaque axiome soit défini par la disposition en question, ni que le sujet serait vis-à-vis de ce contenu dans une relation quelconque de croyance au sens usuel (y compris dispositionnel) du terme (Evans, *ibid.*, 336). Evans note plusieurs différences entre les croyances et les états de connaissance tacite : les premiers sont « au service de multiples projets » et se manifestent de manière holistique, alors que les seconds sont définissables isolément ; les premiers sont soumis à la Contrainte de Généralité (§ 6.5), alors que les seconds sont « inférentiellement isolés » du reste des autres pensées du sujet :

> La possession de la connaissance tacite est exclusivement manifestée dans l'activité de parler et de comprendre un langage ; l'information n'est même pas potentiellement au service d'un autre projet de l'agent, ni ne peut interagir avec d'autres croyances de l'agent... Des concepts tels que ceux que nous utilisons pour la spécifier ne sont pas des concepts dont nous ayons besoin de supposer que le sujet les possède parce que l'état est inférentiellement isolé du reste des autres pensées et croyances du sujet (Evans, *ibid.*, 339)[1].

On ne peut donc pas, de l'aveu d'Evans lui-même, parler ici véritablement d'une *connaissance*.

Nous pouvons donc nous demander si Evans, ou tout autre défenseur du point de vue que l'on a appelé explicatif sur la compétence sémantique, a réellement justifié l'emploi de la notion de connaissance tacite d'une théorie de la signification. Les doutes antiréalistes demeurent entiers : si ce type de connaissance est *par définition* inconscient ou inaccessible, la condition de manifestabilité n'est-elle pas battue en brèche tout autant que quand il s'agit des phrases transcendantes par rapport à la vérification ? Cependant il n'y a rien, en principe, dans la caractérisation de la compétence sémantique comme « tacite » qui doive inquiéter, de prime abord, un antiréaliste. Après tout, Dummett lui-même insiste sur le fait que la compétence sémantique est largement « implicite » (1985), et on pourrait considérer qu'une analyse de cette compétence fondée sur cette notion est *une* manière de satisfaire au réquisit de manifestabilité. La voie serait ouverte, pour l'antiréaliste, d'une analyse de la connaissance implicite des conditions d'assertion des phrases, en termes de capacités pratiques de recognition mises en œuvre. Mais laissons, pour le moment, cette suggestion, pour considérer une autre version du point de vue descriptiviste. Car il y a aussi une autre ligne d'argumentation dans les doutes antiréalistes, plus radicale. Rappelons-nous que Dummett faisait du réquisit de manifestabilité un développement du slogan wittgensteinien : le sens doit être manifesté dans l'usage. Si nous laissons de côté la proposition de Dummett de comprendre ce slogan en termes d'une analyse des conditions d'assertion, la ligne d'argumentation antiréaliste plus radicale peut prendre la forme de ce que Wright (1987 : 23-29) appelle « l'argument de la normativité ». La signification est normative. Connaître la signification d'une expression, c'est savoir comment en évaluer les usages, et savoir quels en sont les usages *corrects*. La question est de savoir comment une théorie de la signification de type vériconditionnel peut représenter ce trait, qu'elle prenne la forme d'une reconstruction rationnelle d'une compétence idéale à la Davidson, ou qu'elle prenne la forme d'une théorie de la connaissance tacite du langage. L'une l'ou l'autre théorie semble conduire à considérer la connaissance des significations comme incorporée dans la théorie qui les représente, soit sous la forme idéale d'une théorie connue par un interprète, soit sous la forme intériorisée d'une compétence tacite. Dans les deux cas, on doit apparemment supposer que les règles sémantiques gouvernant les expressions d'un langage sont représentées dans la théorie. Mais comment la théorie elle-même peut-elle représenter le caractère normatif de ces règles, la manière dont elles gouvernent l'usage et dont les sujets s'y conforment ? Selon la ligne d'argumentation radicale en question, elle ne le peut pas, parce qu'elle assimile maîtrise d'une règle et disposition, norme et fait causal. Toute conception de la compétence sémantique de ce type est vouée à ignorer la normativité de l'usage et des règles. Dans la littérature récente, ce genre d'argument est habituellement présenté à partir des considérations de Wittgenstein sur l'activité de « suivre une règle ».

1. Cf. les critères de la modularité selon Fodor (1983).

7.4. « Suivre une règle » et la connaissance tacite

L'argumentation qui vient d'être esquissée ne figure pas explicitement dans les écrits de Wittgenstein, mais elle est la plupart du temps élaborée

par ses commentateurs à partir d'un certain nombre de thèmes récurrents dans ces écrits. Sans s'engager dans l'exégèse, on peut admettre que ces thèmes sont en général les quatre suivants[1].

(i) Signifier quelque chose par un signe, et suivre une règle, n'est pas un processus ou un état interne au sujet. Cela doit s'entendre en deux sens. En premier lieu, la signification n'est pas constituée par des épisodes mentaux conscients ou « occurrents » (images, idées, expériences). Wittgenstein rejette ainsi toute conception psychologique ou mentaliste de la signification comme processus « privé ». Les épisodes ou processus mentaux peuvent certes accompagner la signification, mais ils ne la constituent pas. En second lieu, la signification n'est pas non plus constituée par des processus mentaux inconscients internes, ni par des mécanismes sous-jacents aux expériences conscientes et au comportement. L'idée commune visée par ces deux thèses négatives est que la signification pourrait être quelque chose d'à la fois interne et parallèle à ce que nous pouvons observer de manière publique et ouverte à tous. (Wittgenstein, 1953 : § 154 : « Essayez de ne pas considérer la compréhension comme un "processus mental" — Car c'est *ceci* qui est l'expression à l'origine de notre confusion. Demandez-vous plutôt : dans quelle sorte de cas et dans quelles sortes de circonstances disons-nous : Maintenant je sais comment continuer », « quand la formule m'est *effectivement* venue à l'esprit », et *ibid.*, p. 218 : « La signification n'est pas un processus qui accompagne un mot. Car aucun *processus* ne pourrait avoir les conséquences de la signification. »)

(ii) Comprendre un signe, et suivre une règle, n'est pas l'interpréter d'une certaine manière. C'est le thème qui figure le plus explicitement dans les sections les plus discutées des considérations de Wittgenstein sur « suivre une règle » (1953 : § 185-219, cf. aussi 1956 : VI, 23-47) Comprendre un signe ne peut pas se produire par l'intermédiaire de ma consultation d'un autre signe, ou d'un état mental qui en constituerait l'interprétation ou la traduction dans un milieu intermédiaire quelconque. Car ce serait alors dire qu'il y a une autre règle qui nous dicte comment nous devons suivre la première, ce qui conduirait à une régression. Mais nous n'interprétons pas une règle, nous la suivons, tout simplement, sans qu'il y ait de notre part un choix quant à la *bonne* manière de le faire. (Cf. 1953, § 198 : « "Mais comment la règle peut-elle me montrer ce que je dois faire à *ce* point ? Quoi que je fasse est, d'après une certaine interprétation, en accord avec la règle." Ce n'est pas ce que nous devrions dire, mais plutôt : toute interprétation flotte encore dans l'air avec ce qu'elle interprète, et ne peut lui donner aucun soutien. Les interprétations par elles-mêmes ne déterminent pas la signification. »)

1. J'emprunte leur exposé ici à Mc Ginn, 1985, et à Wright 1989. Mais on trouverait des caractérisations assez voisines chez Bouveresse (1986), McDowell, 1984, et Hacker et Baker, 1984, 1984b.

(iii) Utiliser un signe en accord avec une règle n'est pas fondé sur des raisons. Quand nous appliquons des mots à des choses, nous ne le faisons pas parce que nous avons des raisons pour penser que c'est la bonne application, ou parce que nous pourrions justifier ou expliquer notre pratique. Quand je suis une règle, je ne consulte pas une sorte de table ou de mode d'emploi qui me dit pourquoi la règle est correcte. Je me contente de la suivre de *cette* manière. (Cf. 1953 : § 211, et § 219 : « J'obéis à la règle *aveuglément*. »)

(iv) Comprendre un signe, c'est avoir la maîtrise d'une certaine technique ou d'une certaine coutume. C'est une pratique. C'est pour ainsi dire la seule « thèse » positive qui paraît se dégager des considérations de Wittgenstein. Il associe couramment la compréhension à une certaine capacité ou propension, qui a un caractère essentiellement *public* et commun. (Cf. 1953, § 150, § 202 : « ..."obéir à une règle" est une pratique. Et *penser* que l'on obéit à une règle n'est pas obéir à une règle. Par conséquent il n'est pas possible d'obéir à une règle "de manière privée" ; autrement penser que l'on obéit à une règle serait la même chose que lui obéir. »)

Aucun de ces thèmes ne peut à proprement parler être considéré comme l'expression d'une thèse ou d'un argument défini. Ici comme ailleurs Wittgenstein semble surtout, négativement, dénoncer certaines images séduisantes, et chercher à les corriger, sans pour autant proposer lui-même une explication de ce que c'est que comprendre une expression ou suivre une règle. Comme je l'expliquerai plus loin, il ne me semble pas qu'il y ait chez lui plus que ces thèmes négatifs. Certains commentateurs, en revanche, ont essayé de donner à ces thèmes une expression beaucoup plus articulée et définie. Les uns, comme Baker et Hacker, ont voulu en tirer une série d'arguments contre les théories systématiques de la signification et la notion de connaissance tacite du langage. D'autres, comme Kripke, ont soutenu qu'elle menaçait l'objectivité même et la réalité de la notion de signification. J'examinerai l'argumentation de Kripke dans la section suivante. Je me concentrerai ici sur celle de Hacker et Baker (1984, 1984a, 1985), qu'on retrouve chez beaucoup d'autres auteurs, mais qui est chez eux la plus polémique. Elle peut se résumer ainsi.

(1) La signification est quelque chose de *normatif*. Connaître la signification d'une expression, et la comprendre, c'est savoir comment évaluer ses usages, et savoir à quelles sortes de conditions ces usages doivent se conformer. Ces conditions sont typiquement appelées des règles. Par conséquent toute analyse de ce que

c'est que « comprendre un langage » doit reposer sur une phénoménologie et une épistémologie adéquate de ce que c'est que « suivre une règle. »
(2) Il y a une certaine conception de la compétence sémantique qui consiste (a) à considérer que cette compétence est une certaine forme de connaissance *théorique*, et (b) à assimiler la compréhension à certains *épisodes, états* ou *processus mentaux internes* du sujet, conscients ou inconscients, et (c) d'après laquelle ces processus *expliquent* la compétence et en cela la *justifient*. L'idée que cette compétence pourrait consister en une connaissance tacite des règles d'une théorie sémantique est l'expression typique de cette conception.
(3) Mais comprendre un langage, et suivre une règle, ne sont pas une forme de connaissance théorique, un *knowing that*, mais des capacités pratiques, des *knowing how*. Traiter la connaissance du langage comme une espèce de connaissance théorique, c'est donc faire une véritable erreur de catégorie.
(4) Comprendre un langage, et suivre une règle, ce n'est pas être dans un état mental particulier. Un tel état, épisode, ou processus, est quelque chose qui est supposé durer ; mais il n'y aucun état isolable, localisé dans une durée, qui pourrait être assimilé à l'activité de « comprendre ».
(5) Il est absurde de supposer que le dévoilement des règles sous-jacentes à la compréhension linguistique, ou à la pratique de suivre une règle, pourrait en quoi que ce soit *expliquer* cette compréhension ou cette pratique. Car si elles consistent en cette pratique, il n'y a aucune raison de supposer qu'il existe un mécanisme ou un processus qui en serait responsable causalement. Les règles que nous pouvons formuler ne sont que des expressions ou des descriptions de la pratique, mais ce ne sont pas des facteurs explicatifs. Toute « explication » de ce genre est donc une pseudo-explication. Si comprendre un langage et suivre une règle devaient jouer un rôle explicatif quelconque, ils devraient constituer des *justifications*. Mais une explication ou une justification doivent s'arrêter quelque part. Or quand nous suivons une règle ou comprenons une expression, nous ne cherchons pas à donner des justifications de ce que nous faisons. Nous agissons, et nous ne consultons pas la règle comme une sorte d'agenda nous dictant ce que nous devons faire.
(6) Le théoricien de la signification est donc victime d'une confusion entre la structure abstraite d'une théorie sémanique et la structure causale et psychologique, et il croit qu'une théorie sémantique ainsi comprise expliquera la compétence. Mais une théorie de la signification n'est pas une théorie ni une explication des règles, pas plus que le système des règles du jeu d'échec n'est la théorie ni l'explication du jeu.

Il est facile de voir en quoi cette argumentation se rattache aux thèmes wittgensteiniens énoncés plus haut (bien qu'il soit plus difficile de voir en quoi elle en découle strictement). Elle est supposée ruiner à la fois la

conception « rationalisante » d'une théorie de la signification comme théorie de la compétence sémantique et la conception psychologique de la connaissance tacite. Selon cette critique, la première conception doit conduire à la seconde. Dans un premier temps, cherchant une explication de la compétence sémantique, le théoricien attribue à un locuteur un ensemble de règles qu'il peut consulter pour justifier sa pratique. Il tombe ainsi dans le Charybde de la confusion entre connaissance théorique et capacité pratique. Dans un second temps, confronté à l'évidente absurdité de l'attribution d'une connaissance consciente des règles linguistiques au locuteur, le théoricien n'a plus d'autre ressource que de tomber dans le Scylla consistant à décréter que cette connaissance dit être inconsciente et tacite. Ce n'est pas moins absurde que la première démarche, car suivre une règle est par définition une activité consciente. Comment un principe peut-il être normatif, régulateur d'une pratique, si les sujets ne peuvent pas connaître ce principe consciemment ni le formuler (Baker et Hacker, 1984 : 312-314) ? Baker et Hacker visent ainsi non seulement Chomsky et ses disciples, mais Davidson, dont la conception rationalisante n'est selon eux qu'une « manière habile » de chercher à éviter d'ouvrir « la boîte de Pandore » que *toute* théorie systématique de la signification qui conférerait à la notion de règle sémantique un statut *descriptif* de la compétence linguistique ne peut manquer d'ouvrir, et également Dummett (*ibid.*, 321-327). Les théories de la signification du type de celles que nous avons examinées dans ce livre sont-elles victimes de cette « mythologie des règles » ? La démythologisation en philosophie est un exercice sain et utile. Mais comme le disait Wittgenstein lui-même, « il y a du charme dans la destruction du préjugé » et les thérapeutiques ne peuvent s'exercer que si elles reposent sur le bon diagnostic et ne soignent pas des maladies qu'elles ont elles-mêmes inventées[1].

Tout d'abord, une théorie de la signification systématique ne repose sur aucune confusion entre une connaissance pratique et une connaissance théorique. Ni Davidson ni Dummett ne nient que comprendre un langage soit essentiellement une capacité pratique. C'est même l'une des thèses

1. Les remarques qui suivent s'inspirent à la fois de Davies, 1986, 1988, et de Wright, 1987, chap. 6, et reprend certains points de Engel, 1985a.

principales de Dummett (cf. les thèses (R) et (C) du § 4.1). Mais d'une part ils distinguent la capacité elle-même de sa *représentation théorique*, qui est ce que vise une théorie de la signification, et d'autre part une chose est de soutenir que la compétence sémantique soit une capacité pratique et autre chose est de soutenir qu'elle soit *intégralement* une capacité pratique. Dummett ne soutient pas cette seconde thèse, mais seulement que la compétence sémantique doit *se manifester* par l'exercice de capacités pratiques. Il cite (1985) le personnage de Wodhouse à qui on demande : « Parlez-vous espagnol ? » et qui répond : « Je n'en sais rien. Je n'ai jamais essayé. » La réponse est absurde, nous dit-il, parce que comprendre un langage repose sur un savoir, qu'on doit avoir acquis d'une manière ou d'une autre (innée ou pas). Le wittgensteinien à la Baker et Hacker entend-il nier ce fait ? Veut-il nier que l'on puisse chercher à expliquer une capacité pratique ? Pourquoi devrions-nous nous contenter de dire que cela fait partie simplement de la « grammaire » du mot « comprendre » que comprendre soit une aptitude, sans chercher à dire en quoi celle-ci peut consister ? En second lieu, le théoricien de la connaissance tacite confond-il structure abstraite et structure psychologique ou causale ? Non, si, comme on l'a vu, on distingue les quatre sens différents de structure sémantique et si on articule correctement la condition du « reflet » proposée par Evans. En troisième lieu, le théoricien la signification est-il voué à confondre l'énoncé de règles sémantiques avec une explication causale, à confondre les règles, qui sont normatives et reposent sur la connaissance de raisons, avec des causes du comportement linguistique ? Aucunement. S'il est indéniable que la recherche d'une base psychologique ou physiologique de la compréhension du langage est bien une tentative d'explication causale, il ne s'ensuit nullement que le théoricien soit par là même en train d'expliquer l'aspect normatif de la compréhension. S'il peut y avoir une explication « cognitive » de la compétence, cette explication ne porte précisément pas sur des notions que l'on ne peut pas articuler consciemment, ni formuler en termes d'états intentionnels ou d'attitudes propositionnelles au sens usuel, comme nous l'avons vu avec Evans. Mais le fait qu'il existe une telle connaissance tacite non propositionnelle n'implique en rien que la signification ne soit pas aussi une notion normative, ni que parler un langage soit une activité gouvernée par des

règles. Le partisan d'une explication de la compétence par des règles connues tacitement doit simplement admettre qu'il parle de règles en un sens distinct du sens ordinaire de cette notion, et que cet usage de la notion répond à des objectifs tout à fait différents de ceux visés quand on veut caractériser la phénoménologie usuelle de la compréhension du langage. Il doit dire qu'il ne cherche pas à rendre compte des raisons ou des justifications que les locuteurs peuvent avoir de parler de telle ou telle manière, et qu'il ne cherche pas à étendre l'espace des raisons : il vise seulement à étendre le domaine de la cognition *au-delà* de l'espace des raisons lui-même (Davies, 1986a). Le problème difficile se réduit alors à ceci : l'emploi de la notion de connaissance tacite est-il légitime ? Le wittgensteinien à la Baker et Hacker soutient que cette notion est incohérente. Mais toute notion de ce genre l'est-elle ? Quand, par exemple, nous attribuons à des locuteurs des intentions de signification, ou que nous disons qu'ils suivent certaines conventions (au sens des théories gricéennes envisagées au chapitre 3), nous ne supposons pas que les locuteurs soient capables d'articuler eux-mêmes ces intentions et ces conventions. Il est essentiel à toute théorie de l'interprétation du discours qui, comme celle que propose Davidson, fait de celle-ci une forme de rationalisation du comportement, que les contenus de pensée attribués ne soient pas nécessairement des contenus que les sujets eux-mêmes soient capables d'articuler. Exiger le contraire irait contre une pratique parfaitement ordinaire d'attributions de pensées, y compris quand nous attribuons des croyances et des désirs à des êtres dénués de langage, commes les petits enfants et les animaux. La question peut certes se poser de savoir si ces attributions sont légitimes, et de savoir ce qu'elles attribuent vraiment (Davidson, 1982), mais nier que la pratique existe et que les attributions portent sur des contenus non nécessairement verbalisables c'est nier l'évidence. Ces remarques ne justifient pas l'emploi de la notion de connaissance tacite au sens « cognitif » envisagé ci-dessus, mais elles montrent qu'il n'y a rien d'*intrinsèquement* confus et mystérieux dans cette notion. Il semble plutôt que ce soit la dénonciation systématique du mystère et de la confusion qui est elle-même mystérieuse et confuse, et que les objections du type de celles de Baker et Hacker ont pour effet de « laisser les choses en l'état ».

7.5. « *Suivre une règle* » *et l'objectivité de la signification*

Des critiques comme celles qui viennent d'être envisagées ne constituent qu'une des versions possibles de l'argument de la « normativité ». Il en existe une autre version, qui ne porte pas tant sur la manière dont on doit concevoir l'activité de suivre une règle ou sur la nature de la compréhension linguistique que, comme je l'ai déjà dit, sur *l'objectivité* et la *réalité* même des règles et de la signification linguistique. Cette critique est donc dirigée contre la conception *réaliste* de l'objectivité de la signification.

Kripke (1981) est l'auteur qui a développé le plus nettement cette ligne de pensée. Selon lui, le fil directeur d'une lecture des remarques de Wittgenstein sur « suivre une règle » peut être fourni par analogie avec la manière dont Hume — du moins selon l'interprétation usuelle — conteste que certains énoncés — en particulier ceux qui concernent nos idées de causalité et les propriétés morales — puissent exprimer des faits réels connaissables et indépendants de nous, et soutient que ces énoncés sont plutôt des *projections* de nos attitudes mentales et de nos réponses affectives, c'est-à-dire rien dont nous aurions une cognition authentique. De même que Hume pose, dans un premier temps, un « problème sceptique » au sujet des notions de causalité et de nécessité, en niant que nous ayons la moindre garantie que le futur ressemblera au passé et que le lien causal corresponde à un *matter of fact*, puis, dans un second temps, propose une « solution sceptique de ces doutes » qui justifie nos croyances causales usuelles par nos habitudes et nos attentes usuelles, Wittgenstein, selon Kripke, pose un problème sceptique mettant en doute la réalité et l'objectivité des règles, puis avance une solution sceptique rendant cette réalité relative à notre accord avec une communauté. Selon Kripke le problème sceptique de Wittgenstein est le suivant. Considérons la classe des faits supposés concernant la signification des expressions, exprimés habituellement quand on dit « Par le terme "X" j'ai toujours, et jusqu'à présent, voulu signifier que *p* ». Supposons donc que l'ensemble de mon comportement et de mes états mentaux ait toujours déterminé, jusqu'à présent, mon jugement que tel ou tel mot s'applique à telle chose, que ce jugement a été correct,

et que les comportements et les états mentaux antérieurs établissent l'existence d'un domaine de faits objectifs et connaissables. L'analogue du problème humien est alors : quelle garantie avons-nous qu'il en est bien ainsi ? Pour reprendre l'exemple fameux de Kripke, quelle garantie avons-nous que, lorsque nous pensons savoir ce que signifie le signe arithmétique « + » et pensons suivre la règle arithmétique usuelle de l'addition, nous connaissons effectivement la signification de ce mot et suivons effectivement la règle ? Pourquoi, par exemple, quand on nous demande d'additionner 67 et 58, pouvons-nous dire que le résultat est 125, et non pas 5 ? Car il serait parfaitement compatible avec tous les faits passés de notre comportement ou de notre psychologie que nous suivions en fait une autre règle, celle de la « quaddition », selon laquelle x *quus* $y = x + y$ *si* $x, y < 57$ *et* $= 5$. Selon Kripke c'est ce « paradoxe sceptique » que Wittgenstein énonce au § 201 des *Recherches philosophiques* :

> C'était là notre paradoxe : aucune conduite ne pourrait être déterminée par la règle, parce que toute conduite peut être rendue conforme à la règle. La réponse était : si tout peut être rendu conforme à la règle, alors tout peut aussi être mis en conflit avec elle. Il n'y aurait aucun accord ou conflit ici.

L'« argument sceptique » consiste à dire que notre comportement, nos dispositions, et nos états mentaux passés sous-déterminent nos jugements sur la signification et les règles que nous suivons. Il n'y a rien, en ce sens, qui nous permette de distinguer ce qui nous « semble correct » de ce qui « *est* correct ». La distinction n'aurait tout simplement pas de sens. La conclusion est radicale : « Il semble que l'idée même de signification s'évanouisse dans les airs » et qu'il n'y ait tout simplement pas de *faits* de signification (Kripke, 1981 : 70-71) La « solution sceptique » que Kripke attribue à Wittgenstein consiste tout d'abord à diagnostiquer la source de la difficulté : elle provient selon lui de l'assimilation des faits de signification à des faits particuliers portant sur nos états mentaux ou notre comportement, et de la thèse qui justifiait notre conception d'un domaine objectif de faits, à savoir l'idée que nos énoncés auraient des conditions de vérité indépendantes. La solution rejette ces assimilations. Kripke suggère parfois, dans un esprit qui paraît se rapprocher de la conception antiréaliste dummettienne, que Wittgenstein

veut nier qu'il y ait plus dans le sens que ce que nous pouvons vérifier, et que le sens de nos énoncés est déterminé par leurs conditions d'assertion plutôt que de vérité (1981 : 73). Mais à la différence de Dummett, il n'associe pas cette idée à un rejet des conclusions de l'argument sceptique. Selon Kripke, la solution sceptique consiste à dire que l'application correcte d'une expression doit être en dernière instance garantie par la communauté dans laquelle il existe un usage ou une pratique de cette expression. C'est cette communauté qui juge de la correction de l'usage. Mais comme chez Hume, cela ne rétablit pas la signification dans ses droits à un statut factuel et connaissable.

L'interprétation de Kripke (« Kripkenstein », comme on l'appelle joliment) soulève trois types de problèmes. Le premier est de savoir si « l'argument sceptique » est cohérent. Le second est de savoir si la « solution sceptique » l'est. Et le troisième est de savoir si argument et solution sont bien ce que Wittgenstein a voulu énoncer dans ses considérations sur les règles. Considérons d'abord le second problème, en supposant pour le moment que l'argument sceptique établit bien ce qu'il prétend établir, à savoir qu'il n'y a pas de faits de signification, ou que « tout langage est dénué de signification » (1981 : 71), ou ce que nous pouvons appeler un « irréalisme » quant à la signification. Il n'y aurait pas de « règles » que nous suivrions quand nous signifions quelque chose par un signe, ni par conséquent quelque chose que nous suivrions. Comme l'a remarqué Wright (1984 : 766-767), c'est une manière curieuse d'établir la conclusion sceptique, puisqu'elle fait appel à la notion même de signification que l'argument est supposé montrer sans application possible. Dans ce cas, l'argument sceptique peut apparaître moins radical qu'il n'en a l'air. Il y a deux options possibles. Il pourrait, par exemple, se ramener à la conclusion que Quine tire de son argument de l'indétermination de la traduction, selon laquelle la notion traditionnelle de signification est inutilisable en droit (bien qu'elle ait une utilité pratique). L'autre option, qui est celle que Kripke lui-même semble adopter, consiste à soutenir une forme d'instrumentalisme, selon lequel la notion de signification n'a aucun contenu factuel, bien qu'elle ait le statut d'une *projection* à partir de nos pratiques courantes d'attribution. Mais si, avec la conception vériconditionnelle, nous admettons qu'il y a au moins un lien nécessaire — même

s'il est non suffisant — entre la signification d'une phrase et ses conditions de vérité, le sceptique de Kripke semble devoir soutenir que nos énoncés portant sur des conditions de vérité n'ont pas eux-mêmes de conditions de vérité, et donc aucun statut factuel. Dans ce cas, comme le dit Wright, comment la distinction entre un discours établissant des faits et un discours purement projectif peut-elle être établie ? Ces difficultés ne sont pas propres à l'analyse de Kripke. Elles sont propres à toute thèse projectiviste en philosophie[1].

Revenons maintenant à la première question : l'argument sceptique est-il défendable ? Est-il correct de dire que ce que nous appelons ordinairement des significations est sous-déterminé par l'ensemble de nos comportements et de nos états mentaux ? Une objection possible serait de dire, selon une inspiration « cartésienne », qu'au moins *certains* états mentaux paraissent pouvoir fixer les significations, à savoir les expériences subjectives et qualitatives que nous associons à des termes désignant des qualités secondes par exemple. Mais ici on se heurtera aux arguments wittgensteiniens qui forment la trame de sa critique du « langage privé » : il est possible que ces expériences *accompagnent* notre compréhension, mais pas qu'elles donnent le *contenu* même des termes en question. Sur ce point, le sceptique kripkensteinien est sur un terrain ferme, du moins si la critique du langage privé est correcte[2]. Kripke lui-même envisage une autre objection : la signification ne peut-elle être fixée par nos *dispositions* au comportement ? Si cette conception est plausible, ou bien on peut soutenir que ces dispositions peuvent déterminer suffisamment la signification, auquel cas on aurait une réponse au scepticisme, ou bien on peut soutenir, avec Quine, que ces dispositions sous-déterminent la signification, auquel cas le scepticisme kripkensteinien se ramène à la thèse

1. Wright, 1984 ; cf. aussi, pour une mise en forme éclairante de cette difficulté, Boghossian, 1989 et 1990. Quand je dis que ces difficultés sont propres à toute thèse projectiviste, cela s'applique également au projectivisme en philosophie morale défendu par Blackburn, 1984, 1993, ainsi qu'aux formes d'irréalisme défendus par Mackie (1977) pour les valeurs morales et par Field (1989) pour les vérités mathématiques.

2. Notons que le point qui vient d'être énoncé ne menace en rien l'accès privilégié que nous pouvons avoir aux significations ni l'asymétrie des attributions à la première personne par rapport aux attributions à la troisième personne. Il nie seulement que ce qui justifie cette asymétrie soit des expériences privées fixant la signification. Cf. ci-dessous.

de l'indétermination de la traduction, et on n'atteint pas une conception plus radicale que cette thèse. Dans cette seconde hypothèse on peut, avec Davidson (§ 2.4), rejeter ces conséquences. Mais Kripke soutient que sa thèse ne se ramène pas à celle de Quine : celle-ci part du point de vue de la troisième personne, alors que la sienne porte sur les attributions de significations que nous faisons à la première personne (1981 : 14-15). Il y a, selon lui, deux raisons de rejeter l'idée que la signification pourrait être déterminée par nos dispositions. La première est que les dispositions sont finies, alors que notre usage des expressions correspond à des capacités infinies. La seconde est que la signification est, à la différence des dispositions, normative : je peux être disposé à utiliser une expression d'une certaine manière, et néanmoins avoir une autre compréhension de cette expression que celle qu'induiraient mes dispositions, alors que je ne peux pas manquer d'avoir une disposition à utiliser une expression en accord avec la manière dont je suis disposé à l'utiliser (1981 : 37). Pourtant il n'est pas certain que les dispositions soient totalement incapables de fixer une signification. Pour prendre un exemple wittgensteinien fameux, si le maçon à qui l'on dit d'ajouter des blocs de pierre à un mur, deux par deux, suivait la règle « déviante » consistant à en ajouter deux jusqu'à ce qu'il ait atteint un total de mille blocs, puis à en ajouter quatre au-delà, il ne pourrait par son comportement manquer de manifester qu'il suit la première ou la seconde règle, s'il se trouvait qu'il ne peut, pour des raisons seulement physiques, porter plus de trois blocs à la fois[1]. En d'autres termes, pour reprendre un point que nous avons rencontré à de nombreuses reprises, nos attributions de dispositions ne sont pas totalement arbitraires : elles peuvent dépendre de nos attributions d'*autres* dispositions et d'autres facteurs, et il est peu plausible dans ces conditions de soutenir qu'il n'y ait *rien* qui puisse fixer l'interprétation.

Mais il y a une considération plus décisive contre l'argument sceptique. C'est qu'il repose sur une conception peu plausible de la manière dont nous nous attribuons des significations à nous-mêmes, par opposition à la manière dont nous attribuons des significations à autrui. Dans ce dernier cas, on peut supposer que les choses se passent comme l'argument

1. Exemple emprunté à Blackburn, 1984.

sceptique le suppose, c'est-à-dire que nous extrapolons à partir des usages précédents d'une expression donnée, de façon *inférentielle* (Wright, 1984 : 774-775). Mais ce n'est pas le cas quand nous nous attribuons la connaissance de significations. Nous faisons, dans ce cas, directement appel à nos intentions ; d'une manière qui n'implique pas une inférence quelconque des cas passés aux cas futurs. Nous avons un « accès privilégié » à nos propres états mentaux que nous n'avons pas quand il s'agit des autres. Or ce trait banal de la « grammaire » des intentions est celui que Wittgenstein est tout à fait d'accord pour considérer comme constitutif (cf. 1953, § 377 : « Quel est le critère de la rougeur d'une image ? Quand c'est l'image de quelqu'un d'autre : ce qu'il dit ou fait. Pour moi, quand c'est mon image, rien »). En ce sens, Wittgenstein ne nie pas qu'il y ait quelque chose comme la signification d'une expression.

D'une façon générale, l'argument sceptique allégué présuppose une certaine conception de l'objectivité de la signification que nous ne sommes pas tenus de considérer comme la seule possible. Il présuppose que le domaine de la signification est constitué par un ensemble de *faits* et que la signification pourrait se réduire à ces faits. Mais sommes-nous obligés d'admettre une telle conception réductionniste en premier lieu ? C'est l'une des implications du holisme de la signification que nous n'avons pas trouvé de raison de rejeter. Si une conception comme celle de Davidson est correcte sur ce point — et j'ai soutenu qu'elle l'était — alors la prémisse principale de l'argument sceptique ne nous est nullement imposée. Une réponse à l'argument sceptique de Kripke peut aussi s'appuyer sur l'analyse de Wright (1989) que nous avons déjà mentionnée au § 5.5, selon laquelle nos jugements linguistiques peuvent, dans des conditions idéales, déterminer l'extension des termes que nous utilisons. C'est l'accord sur ce que nous pouvons tenir comme notre meilleure opinion qui établira les conditions d'application de ces termes. Selon cette conception, conforme à une certaine version du réalisme minimal, les significations sont *moins* objectives que ne le suppose une conception strictement « factuelle », mais *plus* objectives que ne le suppose le scepticisme. La réponse fournie par Peacocke (1992 : 133 sq) et exposée à la fin du chapitre précédent, en termes des « conditions de possession » des concepts, conduirait à une idée voisine : si ces conditions sont spécifiables, alors il y a un

fact of the matter quant à la question de savoir si je suis telle règle ou applique tel concept. Je ne développerai pas ces réponses, et supposerai que le défi sceptique de Kripke peut recevoir une réponse[1].

Enfin, l'interprétation de Kripke est-elle fidèle aux remarques de Wittgenstein lui-même ? Cette question ne m'intéresse pas en elle-même ici, et je n'entreprendrai pas d'y répondre. Mais elle mérite d'être esquissée dans la mesure où elle permet d'envisager une forme d'antiréalisme quant à la signification moins radicale que celle de Kripke[2]. Comme je l'ai déjà remarqué, les quatre thèmes (i)-(iv) sont tous négatifs, et il est peu plausible qu'on puisse en tirer un « argument » et encore moins une « solution ». Ils ne sont pas dirigés contre la réalité même des règles et de la signification, mais contre une certaine image que nous sommes prompts à entretenir à leur sujet. Cette image est celle que Wittgenstein caractérise, aux § 218-219 des *Recherches*, comme la conception d'après laquelle notre compréhension d'une expression (et de toute application d'un concept qui consisterait à faire la « même » chose) seraient déterminées par des règles qui préexisteraient à leur application, et qu'il serait nécessaire de consulter pour les appliquer à des cas particuliers. C'est l'image des règles comme traçant pour nous une « ligne », un « chemin » ou des « rails »[3]. Cette image induit l'idée que suivre la règle doit être une sorte d'acte cognitif ou d'intuition consistant à la saisir (cf. 1953 : § 213).

1. Ce n'est certes pas la seule réponse possible qu'on puisse faire au sceptisme kripkensteinien. Pettit (1991) soutient qu'il repose sur une confusion entre ce que Goodman appelle l'exemplification d'un symbole et son instantiation : la relation d'exemplification entre un symbole et son objet est bien infinie (il y a une infinité de choses qui peuvent l'exemplifier), mais la relation d'instanciation est une relation entre un symbole, un objet et un interprète, et il y a des moyens de déterminer le sens du symbole dans ce cas.

Schiffer (1987 : 173-178) a soutenu que sa propre critique dévastatrice des théories contemporaines de la signification conduit à une « no-theory theory of meaning » qui a des affinités avec le scepticisme de Kripkenstein. J'ai également fait remarquer plus haut (§ 5.5) que sa position a des affinités avec le déflationnisme et le quiétisme quant à la signification. Je n'examine pas ici ses critiques particulières. Mais si les considérations avancées dans ce chapitre sont correctes, elles constituent une réponse partielle à Schiffer, dans la mesure où celui-ci semble se placer dans la perspective réductionniste rejetée ici : il soutient en fait que puisqu'aucune théorie réductionniste ne peut réussir, aucune théorie de la signification ne peut être formulée. J'admets la prémisse, mais je rejette la conclusion. Cf. la conclusion de ce livre.

2. Je m'appuie ici principalement sur Wright, 1989.

3. Cf. 1956 VI, 31 : « A partir du moment où vous avez saisi la règle, vous avez devant vous la route toute tracée. » La métaphore des rails est bien analysée par McDowell, 1981.

Contre cette conception « cognitiviste » ou « platonicienne » des règles (qui fait partie intégrante du platonisme mathématique attaqué dans nombre de ces passages), Wittgenstein préfère dire que se conformer à une règle est une sorte de décision (1953 : § 186, § 213 ; 1956 : VI, 24). Il y a bien ici une forme de scepticisme. Mais il ne porte pas sur la réalité des règles. Il porte sur l'idée qu'en suivant la règle nous suivons pour ainsi dire « à la trace » un ensemble de réquisits indépendants des jugements que nous pouvons articuler. En ce sens, nous n'avons pas d'idée de ce qui constituerait la « direction » que nous impose la règle si nous devions nous en remettre à notre intuition. S'il y a ici un fait qui n'existe pas, ce n'est pas le fait de la règle elle-même, mais le « fait superlatif » (1953 : § 192) en quoi serait supposé consister l'objet de notre « intuition ».

Un corollaire de l'idée selon laquelle nous pourrions consulter la règle par un acte d'intuition est l'idée selon laquelle un certain acte ou processus mental pourrait être responsable de la correction de notre compréhension. C'est ici que Wittgenstein discute ce qui semble être le thème central (iv) : suivre une règle n'est pas l'effet d'une *interprétation*. Ce thème se rattache directement à (i) — suivre une règle n'est pas un processus mental — parce que la manière la plus naturelle de concevoir ce que serait une interprétation de la règle consiste à dire qu'il y a un état mental ou une représentation qui décode la règle. On suppose que l'on a la règle « à l'esprit » sur le modèle de la manière dont on imagine une formule. Mais on peut avoir une telle formule à l'esprit sans savoir ce qu'elle signifie. Et c'est *ici* que l'on rencontre le « paradoxe » que Kripke interprète comme l'expression du « scepticisme » de Wittgenstein : toute interprétation ou formule peut être réconciliée avec la règle. En d'autres termes, une interprétation ne peut m'aider que si elle est *correcte*. Mais c'est circulaire : savoir qu'une interprétation est correcte, c'est savoir précisément quelle règle elle exprime, et savoir comment la suivre. Le passage-pivot des § 198-201 des *Recherches* ne dit pas que du fait que toute interprétation pourrait s'accorder avec la règle il y a une indétermination ou une sous-détermination de la *règle* : il dit que toute interprétation est sous-déterminée *dans l'hypothèse* où suivre la règle est affaire d'interprétation. La plupart des commentateurs ont souligné, contre Kripke, que Wittgenstein n'endossait absolument pas le paradoxe qu'il énonce, mais le

présentait comme une forme de réduction à l'absurde de l'image qu'il attaque.[1] Le sceptique de Kripke met au défi son interlocuteur de dire qu'il sait *quelle* règle son interlocuteur a réellement suivie. Mais le paradoxe de Wittgenstein ne porte que sur la conception interprétative des règles, et n'implique pas qu'on nie qu'il y a une règle à suivre. Wittgenstein le dit même explicitement (1956 : § 197).

Il reste maintenant à voir en quoi peuvent consister les thèses « positives » de Wittgenstein sur les règles et la compréhension. Il est sans doute beaucoup plus difficile de répondre à cette question. Kripke, comme on l'a vu, soutient que la thèse (iv) exprime le caractère essentiellement communautaire de la signification : c'est la communauté qui fixe les normes et constitue la pratique de suivre une règle. Mais cette interprétation est doublement contestable. D'abord elle semble impliquer qu'on ne peut pas avoir de compréhension d'une règle ou d'obéissance à cette règle qui seraient le fait d'un sujet isolé. Mais comme le soulignent (correctement cette fois) Baker et Hacker, ce qui importe, dans l'activité de suivre une règle, n'est pas son aspect collectif, mais son aspect normatif. On ne voit pas en quoi un individu isolé ne pourrait pas suivre une règle[2]. En second lieu, pourquoi la communauté tout entière ne pourrait-elle pas se tromper et suivre des règles déviantes du type de celles envisagées par le sceptique de Kripke sans s'en rendre compte ? La « solution sceptique » de Kripke présuppose que l'accord des membres d'une communauté est une forme de consensus, d'accord statistique fondé sur des jugements communs (et en ce sens les idées de Wright évoquées ci-dessus sont proches). Mais dire cela, selon Hacker et Baker, ce serait retomber dans l'erreur initiale consistant à rechercher, entre la règle et

1. « On peut voir ici qu'il y a une confusion du simple fait que dans le cours de notre argument nous donnons une interprétation après une autre ; comme si chacune d'elles nous contraignait pour un moment, jusqu'à ce que nous pensions à une autre interprétation se tenant derrière elle. Ce que cela montre est qu'il y a une manière de saisir une règle qui n'est pas une interprétation, mais qui est exhibée dans ce que nous appelons « suivre une règle » ou « aller contre la règle » dans les cas réels.

Par suite il y a ici une inclination à dire : toute action qui se fait selon la règle est une interprétation. Mais nous devrions restreindre le terme « interprétation » à la substitution « d'une expression à une autre » (1953 : § 201) ; cf. McDowell, 1984 : 331 ; Mc Ginn, 1985, *passim* ; Hacker et Baker, 1984a : 100 sq. ; Wright, 1984 ; 1989.

2. Cf. Baker et Hacker, 1984a : 41, 71, 81 ; Blackburn, 1984, Goldfarb, 1984.

son interprétation, une sorte de troisième terme qui *expliquerait* leur relation. Mais il n'y a, selon eux, aucune explication de ce genre à donner. Il y a une sorte de relation interne ou d'« harmonie préétablie » (en vertu de nos formes de vie et de notre nature) entre la règle et son application, et par conséquent aucun « fondement » de la règle. Croire le contraire, c'est être victime de l'image même qui est à la source du platonisme quant aux règles. Le sceptique, quand il nie qu'il existe un tel fondement, est en fait victime d'exactement la même image[1].

Wright résume parfaitement, à mon sens, la portée générale des considérations de Wittgenstein sur les règles :

> On nous dit que les réquisits des règles n'existent que dans le cadre d'activités institutionnelles qui dépendent de propensions fondamentales des humains à s'accorder dans leurs jugements ; mais on nous rappelle aussi que de tels réquisits nous fournissent aussi, dans n'importe quel cas particulier, indépendamment de nos jugements, des normes en termes desquelles ces jugements, y compris les jugements consensuels, doivent être évalués. On nous a donc dit ce qui ne constitue pas le réquisit d'une règle dans un cas particulier : il n'est pas constitué par notre accord dans un cas particulier, et il n'est pas constitué, de manière autonome, par une « règle-rail », par rapport à laquelle nous serions dans une relation cognitive inexplicable. Mais on ne nous a pas dit ce qui constitue effectivement ce réquisit ; tout ce que l'on nous a dit est qu'il n'y aurait aucun réquisit sinon pour le phénomène de l'accord dans le jugement.
> Je pense qu'il serait vain de chercher dans les textes de Wittgenstein une suggestion concrète plus positive à propos de cette question constitutive. Sa conception dernière de la philosophie est, bien sûr, conditionnée par une méfiance apparente envers de telles questions constitutives. Ainsi le consensus ne peut pas constituer les réquisits d'une règle parce que nous faisons, à l'occasion, usage de consensus fondés sur l'ignorance ou l'erreur. Mais nous devons nous garder de succomber à notre tendance à ériger à partir d'une pratique où cette distinction reçoit un sens l'image mythologique du suivre-une-règle... L'image mythologique est à l'œuvre dans la philosophie platoniste des mathématiques ; et dans la facilité avec laquelle nous pensons qu'un linguiste privé pourrait établir des critères objectifs de correction et d'incorrection pour lui-même. Il est donc important de l'exposer. Mais une fois exposée, elle n'a pas besoin d'être supplantée. Toute recherche supplémentaire de clarté ne peut être satisfaite que par une sorte d'*übersicht* naturaliste-historique des institutions et des pratiques gouvernées par des règles (1989 : 244).

1. Baker et Hacker, 1984a : 98 sq. Cf. également Bouveresse, 1986 : 50.

Wright appelle cette position « wittgensteinisme officiel ». Elle n'est certes pas sceptique au sens de Kripke : elle ne nie pas qu'il y ait des règles, des significations, et des activités ou des comportements consistant à les suivre ou à les exprimer. Mais elle n'est pas plus hospitalière que le scepticisme kripkensteinien à l'idée d'une théorie systématique de la signification et à l'idée qu'on puisse expliquer la compétence sémantique en des termes psychologiques ou en d'autres termes primitifs. Comme le dit Jacques Bouveresse :

> Supposer que nous devons nécessairement décrire la compréhension d'un langage comme la possession d'un certain type de connaissance qui, du fait du caractère hautement régulier et systématique de l'usage, ne peut être que la connaissance d'une théorie déductive complexe, ce serait, pour un philosophe, se comporter en un certain sens comme le petit peintre Klecksel de Wilhem Busch « qui dessine le profil humain avec deux yeux, parce qu'il sait que l'homme a deux yeux »... Il se pourrait, en outre, que, comme le remarque Putnam, toute connaissance pratique ne soit pas représentable sous la forme d'une théorie (explicite) et que la compétence linguistique soit de ce type. « Nous pouvons acquérir des savoir-faire qui sont trop complexes pour être décrits par une théorie. » Et si nous sommes certains *a priori* que toute aptitude pratique doit pouvoir être représentée de cette façon et qu'elle l'est déjà, d'une manière ou d'une autre, dans quelque « langage » mental ou cérébral, cela signifie simplement que cette façon de voir correspond à ce que Wittgenstein appelle « *eine uns sehr einleuchtende Darstellungsform* » (1981, éd. de 1991 : 61-62).

Bouveresse suggère que Wittgenstein n'entendait sans doute pas nier, à la différence de certains de ses interprètes, qu'on *puisse* donner une telle représentation d'une capacité pratique, ni même qu'on puisse chercher à expliquer la connaissance du langage en partie en ces termes. Mais il aurait sans doute nié que ce genre de représentation et d'explication puisse, et encore moins doit, rendre compte *exhaustivement* de ce que nous appelons « comprendre un langage » et que ce soit le seul mode de description possible. Les règles que nous invoquons dans ce type d'explication, en particulier quand nous parlons de « connaissance tacite des règles » font, comme le dit Bouveresse (*ibid.*), partie de la « préhistoire » du jeu de langage, mais pas de sa pratique. En ce sens, il n'y a de confusion conceptuelle dangereuse que si l'on confond ces deux types de questions en supposant que l'analyse des mécanismes de la compréhension pourra nous

dire en quoi la signification a *pour nous* un tel caractère normatif. On pourrait ainsi réconcilier le partisan de l'approche explicative par la connaissance tacite et le partisan de l'approche descriptive en remarquant qu'ils ne s'occupent pas de la même chose, ni ne répondent à la même question. Le wittgensteinisme officiel les renvoie dos à dos.

La position qu'occupe le wittgensteinisme officiel, dans l'espace logique des diverses analyses que nous avons examinées ici, semble alors être très proche de celle que nous avons appelée au chapitre 5 le *quiétisme*. Le quiétiste quant à la vérité est, rappelons-le, celui pour qui la notion de vérité ne recèle aucune profondeur métaphysique et se réduit à quelques banalités. Le quiétiste quant à la signification est celui pour qui la même chose est vraie de la signification. En ce sens, les critiques qu'adresse Wittgenstein au platonisme des règles, comme celles qu'il adresse au platonisme mathématique, ne sont pas menées au nom d'une *autre* conception métaphysique (un antiréalisme) qui supplanterait la première, mais au nom d'une conception selon laquelle il est illusoire de croire que l'on pourra, à partir de notions comme celle de signification ou de compréhension, aboutir à des conclusions métaphysiques quelconques. La « grammaire » de ces notions est trop variée et trop complexe pour que toute tentative qui consisterait à isoler un sens particulier de ces notions pour en construire une théorie systématique puisse apparaître comme infidèle à la multiplicité intrinsèque de l'*usage*[1].

L'auteur qui s'est fait le plus explicitement l'avocat d'une telle position quiétiste, en s'appuyant en grande partie sur une interprétation des remarques de Wittgenstein, est McDowell (cf. § 5.4). Selon lui, l'erreur que fait un antiréaliste à la Dummett est symétrique inverse de celle que fait le platonicien (le réaliste) quant aux règles et à la signification : alors que ce dernier croit en l'existence de faits objectifs « rigides » de signification gouvernant notre pratique comme des rails, le premier croit que l'objectivité des significations peut être fondée dans des régularités du comportement et des aptitudes pratiques, formulables à partir d'un point de vue « externe » ou « cosmique » sur notre pratique. Mais l'objec-

1. Bouveresse, 1990, a défendu, à mon sens de manière convaincante, ce point de vue au sujet de la philosophie des mathématiques de Wittgenstein, notamment contre l'interprétation intuitionniste et contre la lecture antiréaliste de Dummett.

tivité des significations n'est pas fondée, ni dans des faits transcendants, ni dans une épistémologie particulière de la compréhension. Elle est seulement un donné irréductible de notre appartenance à une communauté, consistant en une capacité intuitive et immédiate exhibant des régularités et des normes qui sont directement à notre disposition, « pour ceux qui ont des yeux pour voir » (McDowell, 1987). Selon McDowell, cette conception, qui ne peut être autre que « modeste », doit présupposer, dans l'énoncé de ce que signifient les mots de notre langage, que nous connaissons déjà leurs significations. Toute autre conception tombe dans les impasses et les illusions propres au réalisme ou à l'antiréalisme en tant que thèses métaphysiques et explicatives. Elle garantit à la fois la manifestabilité du sens, puisqu'il est « là », à la surface même de notre pratique, du fait que nous partageons un langage public, et l'objectivité du sens, puisqu'il est constitué par des normes qui sont indépendantes de nous. Tout ce qui nous reste à faire est de décrire notre pratique, de l'intérieur de notre jeu de langage et de nos formes de vie[1].

Tout comme j'ai cherché, au chapitre 5, des raisons de résister au charme discret du quiétisme quant à la métaphysique de la signification, je chercherai ici à résister à ce charme quant à l'épistémologie de la compréhension. Au wittgensteinien officiel et au quiétiste à la McDowell, on peut poser les questions que posait Gareth Evans :

> Soutiennent-ils que toutes les capacités à comprendre des phrases nouvelles (à savoir, j'y insiste, ce qu'elles signifient) sont également inexplicables, ou croient-ils qu'il y a encore une place pour un type d'explication de ce qu'un locuteur sait ce qu'une nouvelle phrase veut dire quand et seulement quand on peut montrer qu'elle contient des éléments qui figurent aussi dans des phrases dont l'usage est déjà familier ?
> Et dans l'hypothèse où leur réponse à cette première question est oui, j'en pose une seconde :
> Croient-ils qu'il est suffisant de donner une explication du type de celle que l'occurrence d'expressions familières rend possible simplement en montrant que la phrase nouvelle contient bien des expressions qui figurent aussi dans des phrases dont l'usage est déjà familier, ou croient-ils, étant donné la possibilité évidente d'ambiguïtés, que l'on doit donner une autre explication ? Si c'est le cas, est-ce que cette autre explication diffère d'un énoncé de la régularité entre l'usage ancien et l'usage nouveau ? (Evans 1981 (1985) : 341-342.)

1. Il y a des similarités, qu'indique McDowell lui-même (1977), entre cette position et certaines théories romantiques du langage comme celles de Herder, de Humbolt et de Hegel. Je les ai analysées dans Engel, 1990.

Comme je l'ai dit plus haut, le wittgensteinien n'est pas obligé d'adopter une position aussi radicale que celle à laquelle conduit une réponse positive à la première question : il peut soutenir qu'une théorie systématique de la signification et de la connaissane tacite du langage peuvent expliquer *en partie* les aptitudes compositionnelles et la productivité en question, mais que l'on n'aura pas pour autant rendu compte de la normativité de la signification. Mais s'il adopte un quiétisme radical comme celui de McDowell, il est difficile de voir comment il peut éviter de répondre positivement aux deux questions posées par Evans. En ce cas, il doit nier qu'une théorie de la signification, sous quelque forme que ce soit, puisse expliquer quoi que ce soit, et qu'il y ait quoi que ce soit à expliquer. Il doit soutenir que le projet même n'a aucun sens. Cette position me paraît confiner à l'obscurantisme et à ce que Dummett, dans sa réponse à McDowell (1987 : 268), appelle une « mystification » : « Si les enfants vous demandent du pain, leur donnerez vous une pierre ? »

7.6. Sémantique et psychologie

J'en conclus qu'une position qui réduirait la tâche d'une analyse de la compréhension linguistique à un pur point de vue descriptif ou modeste au sens de McDowell a peu de chances de satisfaire aux réquisits mêmes que l'on peut tenir comme plausibles quand on entreprend une telle analyse. Cela n'exclut pas, en principe, la possibilité d'une conception descriptive comme celle qu'envisage Dummett, que nous avons caractérisée au § 7.1 comme une tentative pour rendre compte *à la fois* des traits systématiques de la compréhension linguistique et les faits relatifs à l'usage et à la pratique. Mais à la réflexion, la difficulté mise en avant par Evans — qu'une pure théorie de l'usage ne permet pas d'expliquer le caractère systématique de notre compréhension — affecte aussi la position de Dummett. Bien que cette position se veuille molécullariste, elle ne nous permet pas, dans l'hypothèse où les produits d'une théorie systématique de la

signification sont rattachés à des « aptitudes pratiques » manifestées dans l'usage, de voir en quoi les aptitudes pratiques en question sont fondées dans des états psychologiques du locuteur. Dummett admet bien que la compétence sémantique est fondée dans des dispositions ou des capacités qui font de sa connaissance une connaissance largement implicite, mais il ne nous dit pas *comment* celles-ci sont reliées à ce que le locuteur *sait*, quand il connaît le sens d'une expression, ce que les théorèmes d'une théorie systématique de la signification énoncent (Smith 1992 : 130). De ce point de vue, la difficulté est la même que celle qui affecte la position davidsonienne « officielle », qui laisse les bases psychologiques de la compréhension indéterminées. Nous avons vu également que si une théorie de la signification antiréaliste doit prendre la forme d'une théorie des conditions d'assertion, le réquisit de compositionnalité est difficile à satisfaire (§ 5.2).

Il me semble alors nécessaire d'admettre qu'une théorie de la compréhension doit incorporer une théorie de la connaissance tacite, et doit répondre aux exigences du point de vue explicatif. Mais ce n'est pas chose facile. Revenons à la conception d'Evans. Il propose, on s'en souvient, que l'on assimile la connaissance tacite qu'a un locuteur d'un langage simple L contenant 100 phrases, avec la possibilité d'attribuer à ce locuteur vingt dispositions distinctes, correspondant chacune aux axiomes d'une théorie sémantique compositionnelle T2. Une telle théorie décrirait, conformément à la condition du reflet, *à la fois* la structure sémantique pertinente de L *et* les états causaux responsables de la compétence psychologique du locuteur. Ce qui justifie, dans ce cas précis, l'ascription d'une connaissance tacite de T2 plutôt que de T1 (qui énumère simplement les phrases-T correspondant aux cent phrases de L) est le fait que dans T2, mais pas dans T1, les dérivations des théorèmes spécifiant la signification de phrases *Fa, Fb, Fc,* etc., impliquent un facteur commun, à savoir l'axiome pour « *F* », et pareillement, pour les phrases *Fa, Ga, Ha,* etc., l'axiome commun pour « *a* » — axiomes qui expriment dans chaque cas une « capacité causale » supposée repérable empiriquement. Mais si l'on voit bien en quoi cette suggestion est conforme à la condition du reflet, on ne voit pas en quoi elle justifie cette condition. Comme l'a remarqué Wright (1987 : 231), pourquoi adopter T2 ou une théorie quelconque dont les

axiomes sont *sémantiques* si notre but est de refléter la structure *causale* des dispositions correspondant aux théorèmes établissant la signification ? Pourquoi une structure causale devrait-elle être articulée dans une structure sémantique ? La question se pose d'autant plus que l'analyse de Evans manifeste un trait curieux, noté par Wright (*ibid*) : il nous dit que la connaissance tacite de T2 met en jeu 20 états du sujet, alors que le nombre des axiomes de T2 est de 21. Le problème n'est pas que le nombre des axiomes ne coïncide pas avec celui des dispositions, puisque Evans précise bien qu'il y a une disposition pour chaque *expression*, et pas pour chaque *axiome*. Le problème est plutôt qu'un sujet qui aura les 20 dispositions interconnectées — une pour chaque axiome relatif à une expression — sera, par hypothèse, disposé à assigner des significations correctes aux phrases de L. Mais comme le dit Wright, T2 serait boiteuse sans l'axiome compositionnel, bien que celui-ci soit, du point de vue des dispositions du sujet, redondant. Wright (*ibid.*) soutient qu'Evans se trouve face à un dilemme : d'un côté l'axiome compositionnel est superflu quand on considère les dispositions, et de l'autre il est essentiel à une théorie sémantique compositionnelle. Cela montre, selon Wright, que la correspondance alléguée entre structure causale et structure sémantique est douteuse.

Il est certain qu'une théorie compositionnelle doit articuler une structure similaire à celle de T2. Mais nous pouvons échapper à la difficulté créée par la différence entre le nombre des axiomes et celui des dispositions en formulant une troisième théorie T3, qui contiendrait, comme T2, un axiome distinct pour chaque nom, mais qui au lieu des axiomes de T2 pour les prédicats ait 10 axiomes du type suivant :

> Une phrase couplant un nom au prédicat *F* est vraie ssi l'objet dénoté par le nom est chauve

On incorporerait ainsi l'axiome compositionnel dans les axiomes pour les prédicats. T3 aurait alors 20 axiomes, et serait extensionnellement équivalente à T1 et à T2. La difficulté soulevée par Wright se ramènerait alors au problème posé posé par Quine, c'est-à-dire au fait que nous n'aurions pas de raison d'attribuer la connaissance tacite de T2 plutôt que de T3. Nous devons donc admettre que la connaissance tacite est relativement

indéterminée. Mais elle ne l'est que *relativement*, parce que s'il est clair que T2 et T3 ne sont pas logiquement équivalentes, elles sont extensionnellement équivalentes, en ce sens qu'elles permettent d'attribuer à un locuteur la même capacité à discerner la structure de L (ce qui n'était pas le cas pour T1 *vs* T2). Bien que les algorithmes soient distincts, il y a pourtant un sens où c'est la *même* information qu'ils calculent (Davies, 1987).

On répond ainsi à la première objection de Wright (pourquoi 21 axiomes pour 20 états de connaissance tacite ?). Mais on ne répond pas à son autre objection : pourquoi une structure *causale* devrait-elle être représentée par une structure *sémantique* ? Wright (1987 : 235) suggère qu'il n'y a aucune raison de supposer qu'il y ait un lien entre un niveau proprement sémantique d'explication et un niveau causal, car les états causaux pourraient n'avoir aucun lien avec les régularités exhibées au niveau sémantique. Si l'on avait, par exemple, à articuler la « connaissance tacite » qu'a un pigeon voyageur de diverses localisations, nous pourrions articuler une telle structure causale ; mais nous n'aurions aucune raison de la mettre en parallèle avec une structure sémantique et compositionnelle. Il suffirait de calculer l'information spécifiée par un certain mécanisme, et de spécifier la manière dont elle est réalisée dans l'appareillage sensoriel et physique de l'oiseau. C'est un trait purement contingent aux humains que l'information calculée par un certain mécanisme puisse être mise en corrélation avec une structure sémantique. Il n'y a aucune raison de supposer que cette corrélation soit de principe.

La proposition de Wright appelle immédiatement une comparaison entre le problème qui nous occupe ici, celui du type d'explication fourni par une théorie sémantique, et le problème familier de l'explication en psychologie cognitive. Tel qu'il a été posé notamment par David Marr dans ses travaux sur la vision[1], une explication cognitive repose sur une distinction entre trois niveaux : 1 / un premier niveau abstrait ou *computationnel*, qui spécifie quelle sorte de fonction un certain système cognitif calcule, 2 / un second niveau *algorithmique* spécifiant la nature des processus computationnels mis en jeu, et 3 / un troisième niveau, *physique*,

1. Marr, 1982. Cf. ma discussion de ces problèmes dans Engel, 1992.

qui indique comment l'algorithme en question est réalisé dans la composition matérielle ou physique du système. Dans le cas où l'information calculée est une information sémantique, le niveau computationnel supérieur de Marr est celui de la théorie sémantique qui caractérise l'information calculée par un sujet, dans les conditions idéales, le niveau algorithmique est constitué par la détermination des dispositions causales associées à chaque expression, et le niveau physique par la spécification de la base neurophysiologique de ces dispositions. La question posée par Wright est celle de savoir pourquoi les spécifications du niveau supérieur devraient nécessairement refléter celles du niveau inférieur physique et celles du niveau dispositionnel, où les structures proprement causales opèrent. C'est une question profonde et difficile, qui divise les théoriciens des sciences cognitives. Elle porte sur la légitimité de la démarche préconisée par Marr, consistant à partir du niveau supérieur, pour atteindre, en descendant vers les niveaux inférieurs, le niveau physique. Dans les modèles dits « classiques » en sciences cognitives, cette difficulté est celle de savoir comment on passe d'un niveau où l'information est codée sémantiquement à un niveau où elle est réalisée physiquement. Ces modèles résolvent la difficulté en supposant qu'il existe un niveau intermédiaire de symboles, un langage de la pensée où des symboles physiques entrent dans des relations computationnelles comme dans les ordinateurs (Fodor 1975, 1987). Mais comment passe-t-on du sémantique au symbolique et de là au physique ? Ce sont précisément sur ces articulations de niveaux que s'opposent les modèles classiques et les modèles connexionnistes[1]. Mais nous pouvons prendre la mesure de la difficulté sans entrer ici dans ces débats. Nous avons, en présentant la théorie d'Evans, vu que ce dernier refusait d'assimiler les états de connaissance tacite à des états intentionnels ordinaires d'attitudes propositionnelles. Suivant une terminologie due à Stich (1978), nous pouvons dire que ce sont des états « subdoxastiques ». A la différence des états d'attitudes propositionnelles, qui sont holistiques, qui manifestent une « promiscuité inférentielle », les états subdoxastiques sont informationnellement et inférentiellement isolés. Cette différence semble tenir au fait que les premiers sont des états doués effectivement d'un contenu

1. Cf. Fodor-Pylyshyn, 1988, Engel, 1992, chap. 3.

sémantique, alors que les seconds ne véhiculent pas un tel contenu, bien qu'il soit légitime de dire qu'ils véhiculent une certaine *information*. La question de Wright peut être posée ainsi : si la seule contribution qu'apporte une théorie sémantique à une psychologie de la compétence sémantique consiste à spécifier l'information sémantique au niveau abstrait ou « computationnel », elle ne nous dit rien, par elle-même, des états cognitifs sous-jacents des locuteurs. Il est donc impossible d'attribuer une connaissance tacite des axiomes d'une théorie de la signification au titre d'un état quelconque d'un locuteur réel. Tout ce qu'il est possible de dire est qu'une théorie sémantique éclaire ce qu'un locuteur *devrait* savoir s'il devait parler un langage faisant l'objet de la théorie sémantique en question. On en reste donc à la conception « rationalisante » de la structure sémantique — notre seconde notion de structure ci-dessus § 7.2. Mais si nous voulons passer à un niveau inférieur cognitif, et appeler connaissance tacite le contenu d'information véhiculé par les états subdoxastiques, nous n'avons aucune garantie que *cette* information reflète ou corresponde à celle que spécifie la théorie sémantique au niveau abstrait. Les deux types d'information, l'information sémantique telle que nous pouvons l'attribuer à des locuteurs d'un point de vue interprétatif, et l'information « cognitive » telle que nous pouvons l'attribuer du point de vue explicatif, semblent de nature essentiellement différente, ou tout au moins il n'y aucune garantie qu'elles coïncident. En d'autres termes, il n'y a aucune garantie que ce que nous exprimons dans notre langage quand nous attribuons des contenus propositionnels ordinaires et quand nous spécifions la contribution d'une expression au sens d'une expression totale, soit ce que les états cognitifs sous-jacents véhiculent. Supposons que nous connaissions cette dernière information : comment allons-nous la délivrer ? Comment va-t-elle délivrer au sujet lui-même, à la première personne, la connaissance qu'elle est supposée contenir au niveau infrapersonnel ?

Ce problème paraît insoluble si l'on part du principe que l'information donnée par une théorie sémantique est pleinement intentionnelle et sémantique, alors que l'information véhiculée par les états subdoxastiques ne l'est pas. Le dilemme, déjà rencontré, auquel on est confronté est qu'ou bien on stipule que les deux types de « connaissance » n'ont rien en

commun, auquel cas on ne voit plus bien en quel sens parler de connaissance tacite de *contenus* déterminés, ou bien on stipule qu'elles ont quelque chose en commun, auquel cas la connaissance tacite en question risque d'être triviale — une simple projection gratuite, dans des structures causales, de structures sémantiques. Mais il y a une réponse possible, qui a été proposée par Peacocke (1986a) et Davies (1987). Revenons au constat énoncé plus haut : l'attribution de connaissance tacite à un locuteur est relativement indéterminée. Mais ce fait ne va pas nécessairement à l'encontre de l'attribution de connaissance tacite. On peut soutenir en effet que tout comme deux algorithmes distincts peuvent calculer la même fonction, deux algorithmes différents peuvent véhiculer la même information. C'est ce qui se produit avec T2 et T3, où il est clair que la composante commune est l'information contenue dans l'axiome compositionnel. Peacocke (1986a) a soutenu que cet élément informationnel commun, qui rend compte, dans l'analyse de Evans, du fait qu'un sujet soit capable de comprendre toutes les phrases contenant un nom et un prédicat, n'est pas représenté au niveau 1 computationnel de Marr. En effet, si l'on compare cette fois T2 à T1 (la théorie prenant la forme d'une liste d'axiomes pour chaque phrase de L), c'est la *même* fonction, définie en extension, qui se trouve spécifiée pour le locuteur de L opérant d'après T1 ou d'après T2. En d'autres termes, deux locuteurs qui connaîtraient tacitement respectivement T1 et T2 parviendraient à déterminer également la signification des phrases de L. Le niveau computationnel ne peut donc pas représenter la différence entre les deux compétences (l'une est structurée, l'autre pas). Or quand nous spécifions l'information sur laquelle reposent les mécanismes représentés cette fois par T2 et par T3, opérons-nous au niveau 2 de Marr (algorithmique) ? Non, car plusieurs algorithmes sont compatibles avec la compétence structurée des locuteurs. Il est clair que nous ne sommes pas non plus au niveau inférieur physique 3. Peacocke propose donc que le type d'information spécifié par T2 et T3, ou toute autre théorie qui rendrait compte identiquement de la structure des phrases, soit caractérisé comme relevant d'un niveau *intermédiaire* d'explication situé entre le niveau 1 et le niveau 2. Il l'appelle « niveau 1,5 » et donne d'autres exemples d'explications psychologiques cognitives situées à ce niveau. C'est à ce niveau intermédiaire qu'opère une théorie

cognitive si elle fournit l'information, ou le contenu, sur lesquels reposent deux algorithmes différents. Mais cette information n'est pas seulement abstraite, comme celle spécifiée au niveau I : elle a une influence causale sur la capacité du locuteur à reconnaître la signification des phrases.

Peacocke ne fait, en réalité, que reformuler l'analyse de Evans, et se heurte aux mêmes difficultés : du fait que deux algorithmes « reposent » sur la même information, il ne s'ensuit pas que cette information soit *utilisée* par le sujet et qu'elle ait une pertinence causale sur sa compétence, et la nature des algorithmes en jeu n'est pas totalement spécifiée, puisqu'il y a indétermination entre T2 et T3. Mais rien ne nous interdit de dire que l'information n'a pas de répondants causaux. Rien ne nous y autorise non plus, il est vrai, et en ce sens le problème reste entier. Mais nous avons au moins une réponse claire à notre question initiale : comment une théorie sémantique peut-elle faire partie d'une explication psychologique et d'une explication causale de la compétence sémantique ? La réponse est que

> Une théorie sémantique peut faire partie d'une véritable explication psychologique de la capacité qu'a un sujet de comprendre des phrases, même si le sujet ne construit pas, à un niveau subpersonnel, des dérivations à partir de cette théorie, et même si la théorie n'est pas explicitement représentée mentalement chez le sujet. La théorie sémantique peut faire partie d'une explication psychologique en spécifiant l'information sur laquelle repose une certaine computation ou un certain mécanisme (Peacocke, 1986b : 115).

Cette réponse ne s'identifie pas à celle selon laquelle une théorie sémantique contribuerait à une explication psychologique de la compétence seulement en spécifiant une information abstraite établissant seulement ce que le système cognitif a à calculer, et laissant les détails de l'explication au psycholinguiste et aux neuropsychologues. Elle suppose qu'il existe une forme de contenu, distincte du contenu intentionnel ou personnel, impliquée au niveau des computations subpersonnelles du système cognitif, et que la notion même de « computation » doit être associée à des contenus, et non pas être simplement associée à des mécanismes causaux et à des processus neurophysiologiques. C'est une hypothèse, mais elle est cohérente. Seule une conception qui assimile d'entrée de jeu tout contenu séman-

tique à un contenu intentionnel, accessible à la première personne et interprétable en termes de la psychologie usuelle des attitudes propositionnelles peut rejeter ce genre d'hypothèse. Quelle raisons avons-nous d'adopter une telle conception ? Celles qui ont été analysées dans ce chapitre : notre notion ordinaire de contenu, de signification est normative, et elle suppose que les agents aient accès à la première personne, et, à travers la communication avec autrui, à leurs contenus de pensée et de signification. Il ne s'agit pas de nier que ce que nous appelons « signification », « pensée » ou « compréhension » repose sur de tels critères. Mais cela interdit-il d'envisager une explication de la compétence sémantique à un niveau subpersonnel et une notion de contenu sémantique qui ne réponde pas au caractère normatif des significations au sens usuel ? Je n'ai pas trouvé de raison convaincante de le faire.

Revenons aux trois points de vue, interprétatif, descriptif et explicatif, proposés ci-dessus. Le point de vue interprétatif est fidèle au caractère de nos attributions de signification à la troisième personne. Mais il ne nous donne aucune prise sur une explication psychologique possible de la compétence sémantique. Le point de vue descriptif est fidèle à la phénoménologie de la compréhension du langage, au fait qu'il n'y a, en un sens, rien de plus dans ce que signifient nos mots que ce *nous* comprenons quand nous les utilisons. Mais il menace de glisser vers une forme de quiétisme qui renonce à toute possibilité même d'expliquer le phénomène en question. Le point de vue explicatif promet une explication de ce genre, mais il reste insensible au caractère normatif de la signification. Nous pouvons nous demander si ces divers points de vue s'adressent au même type de faits, et si ce n'est pas une erreur de catégorie que d'aborder les uns avec les concepts qui semblent appropriés à d'autres. Mais nous pouvons aussi considérer qu'ils offrent des points de vue distincts, à des niveaux distincts, d'un seul et même ensemble de faits, complexes mais néanmoins liés entre eux. Je n'ai pas montré comment c'est possible, mais cela me paraît au moins plausible.

Conclusion

Only connect.
E. M. Foster.

Dans son formidable livre *Remnants of Meaning* (1987 : 1-17), Stephen Schiffer envisage un philosophe — peut-être lui-même dans sa jeunesse — qui soutiendrait les neuf thèses suivantes.

(1) *Il y a des faits sémantiques* : certains sons et marques ont une signification, certains sont vrais, certains ont une référence.
(2) *Toute langue naturelle a une théorie compositionnelle de la signification* : une théorie finitaire, qui spécifie la signification de toute expression de la langue.
(3) *Toute théorie compositionnelle de la signification détermine une condition de vérité pour chaque phrase du langage.*
(4) *On ne pourrait pas rendre compte de la capacité qu'ont les locuteurs de comprendre des phrases nouvelles si l'on n'admettait pas* (3).
(5) *Il y a des faits intentionnels* : les humains ont des croyances, des désirs et d'autres attitudes propositionnelles douées de contenu.
(6) *Les croyances et autres attitudes sont token-token identiques à des états neurophysiologiques des humains.*
(7) *Les croyances sont relationnelles* : croire que *p* est une certaine relation d'un individu à une certaine entité.
(8) *Les faits sémantiques et psychologiques ne sont pas irréductiblement tels, mais peuvent être exprimés en termes physiques, sans faire appel à des notions sémantiques ou psychologiques.*
(9) *Les faits sémantiques se réduisent à des faits psychologiques, au sens où l'a envisagé Grice.*

Schiffer entreprend de montrer que toutes ces thèses sont fausses à l'exception de (4) et de (6), et en conclut qu'on doit renoncer à toute théorie

positive de la signification et des contenus mentaux, et adopter ce qu'il appelle la réponse « défaitiste » consistant à souscrire à une « *no-theory theory of meaning* ». Il voit certaines affinités entre cette réponse et le scepticisme quant à la signification de Kripke (*ibid.* : 173-178), bien qu'il admette avec le sens commun la vérité de (1) et de (5), mais en un sens seulement « trivial » et « philosophiquement non intéressant ». Je n'ai pas examiné dans ce livre toutes les thèses incluses dans la liste de Schiffer, ni les raisons pour lesquelles il les critique, bien que nous ayons rencontré à de nombreuses reprises des objections très voisines des siennes. Je n'ai pas discuté la thèse physicaliste (6) ni le monisme anomal de Davidson et ai seulement indiqué en quoi il se rattache à ses thèses en philosophie du langage. Je n'ai pas envisagé la critique que mène Schiffer contre (2)-(3) à partir de sa critique de (7), la théorie relationnelle des atttitudes propositionnelles. Schiffer a aussi des raisons indépendantes (1987 : chap. 7) de nier qu'une théorie sémantique doive être compositionnelle. Ces raisons ne me paraissent pas convaincantes, et je me bornerai ici à répéter, comme au § 6.5, que la compositionnalité est une caractéristique essentielle des significations et des contenus mentaux conceptuels. Ici la nomenclature de Schiffer nous servira seulement à récapituler les conclusions auxquelles nous sommes parvenus. Elles diffèrent assez de celles auxquelles lui-même parvient.

Les thèses (2)-(3) (4) peuvent être attribuées à Davidson, et forment, comme on l'a vu, la base de sa théorie de la signification. Davidson soutient aussi une version de la thèse (7), selon laquelle les croyances sont des relations à des phrases. Il admet aussi, en vertu de son monisme anomal, la thèse (6). Mais il rejette nettement les thèses (8) et (9), parce qu'il rejette toute théorie réductionniste de la signification et de l'intentionnalité. Le statut de (1) et de (5) est plus problématique chez lui.

Davidson ne soutient pas les thèses (1)-(4) sans les assortir de conditions particulières. Il défend l'idée qu'une théorie sémantique compositionnelle et véricondi tionnelle doit prendre la forme d'une théorie-T à la Tarski, et qu'elle ne peut faire office de théorie de la signification pour une langue naturelle que si elle satisfait aussi à des conditions formelles et empiriques bien précises. La question qui nous a occupés dans la première partie de ce livre a été celle de savoir si ces conditions permettaient bien

à Davidson de justifier cette proposition. Notre réponse a été mitigée. Nous avons admis que quand il s'agit de décrire la structure compositionnelle effective d'une langue naturelle, les contraintes d'extensionnalité et de forme logique de Davidson ne s'imposaient pas nécessairement, et qu'elles ne se justifiaient que dans la mesure où l'on admet le cadre de la théorie de l'interprétation radicale. Le poids du programme sémantique repose alors sur cette théorie. Nous avons alors rencontré trois ensembles de difficultés.

Le premier porte sur la possibilité, à partir d'une théorie de l'interprétation, de rendre compte des usages pragmatiques du langage, en particulier dans l'hypothèse, qui est celle de Grice, où l'interprétation des énoncés doit faire appel aux intentions de signification des locuteurs. Nous avons vu que rien, dans la théorie de l'interprétation davidsonienne, n'empêche de rendre compte de ces usages, à condition cependant qu'on accepte son analyse sémantique des modes, et qu'on renonce à interpréter le programme de Grice comme un programme réductionniste, au sens des thèses (8) et (9) de Schiffer. A nouveau, le poids de l'argument repose ici sur l'adoption de la théorie de l'interprétation radicale.

Le second ensemble de difficultés porte sur les doutes émis par l'antiréalisme quant à la capacité, pour une conception véri conditionnelle, d'être une authentique théorie de la signification, à partir du moment où cette conception paraît postuler l'existence de conditions de vérité transcendantes par rapport à l'usage des phrases et à leur vérification possible. J'ai examiné trois volets de cette objection. 1 / Tout d'abord, un réaliste quant à la signification est-il commis à l'image de conditions de vérité indétectables et non manifestables que critique l'antiréaliste ? J'ai soutenu que non. Dans la version holistique d'une théorie de la signification de Davidson, la connaissance qu'ont les locuteurs des significations n'est pas « non manifestable » : elle se manifeste dans les conditions propres à l'interprétation, et une version plus forte du réquisit de manifestabilité, celle que défend Dummett, doit se heurter au caractère holistique de toute attribution de signification. Exiger plus, et supposer que cette connaissance se manifeste dans des capacités pratiques à utiliser les phrases, c'est ignorer ce caractère, et présupposer la thèse molécula riste qu'il faudrait établir. Mais cet argument contre l'antiréalisme présuppose lui aussi quelque

chose : qu'une forme de holisme de la signification est légitime. 2 / Ensuite, la conception vériconditionnelle implique-t-elle la conception d'une vérité transcendante par rapport à toute vérification ? J'ai soutenu que non. Elle revient à admettre une série de platitudes qui tiennent à l'usage décitationnel du prédicat de vérité et à sa relation à l'assertion, mais elle n'implique pas une forme de réalisme « externe » qui ferait de la vérité une propriété indétectable de nos énoncés. Je n'en ai pas pour autant conclu, comme les partisans d'une conception déflationniste ou quiétiste, que le débat réalisme/antiréalisme se vidait de toute substance. C'est pourquoi j'ai formulé une position, le réalisme minimal, qui admet les contraintes épistémiques générales de la manifestabilité de la signification, qui rejette l'idée d'une vérité transcendante, mais qui ne réduit pas la vérité à l'assertabilité. J'ai soutenu qu'il y avait plusieurs versions possibles de cette thèse et que le réalisme de Davidson était l'une d'elles. Mais ici encore, le poids de l'argumentation repose sur le holisme. C'est à ce point que j'ai pris mes distances par rapport à Davidson. Bien que ce holisme soit parfaitement correct au plan méthodologique, il impose des conditions trop lâches sur l'individuation des contenus intentionnels et des significations. Il n'est pas clair qu'un réalisme minimal doive prendre la forme d'une théorie de l'interprétation au sens de Davidson. En ce sens, je crois possible, comme Peacocke, d'adopter certains réquisits épistémiques de l'antiréalisme sans souscrire pour autant au vérificationnisme, et d'adopter un holisme minimal sans souscrire au holisme radical. 3 / Enfin, une conception vériconditionnelle peut-elle rendre compte de la compétence effective des locuteurs ? Ici l'antiréaliste avance sur un terrain miné : s'il adopte une version trop forte de la thèse selon laquelle il n'y a rien de plus dans la signification que ce que *nous* pouvons comprendre, il risque de tomber dans une forme de scepticisme radical qui menacerait toute possibilité de construire une théorie objective de la signification. Mais le réaliste davidsonien court un risque symétrique : s'il rejette toute explication possible de la compétence linguistique en termes psychologiques ou « cognitifs », au bénéfice d'une conception seulement idéale de cette compétence, en insistant sur le fait que la signification est une notion irréductiblement normative, il risque de tomber dans une forme de quiétisme comme celle de McDowell (et peut-être comme celle de

Wittgenstein) d'après laquelle la compétence sémantique est seulement un fait qu'il n'est pas possible d'expliquer autrement qu'en systématisant notre pratique d'interprétation. Il n'est pas facile de déterminer exactement où se situe, dans cette géographie doctrinale complexe, Davidson lui-même. Davidson ne propose sans doute pas la même conception de la normativité de la signification que celle du wittgensteinien : il soutient qu'elle dérive des principes normatifs qui règlent *a priori* toute attribution de significations et d'états intentionnels, alors que ce dernier soutient sans doute qu'elle dérive de notre « pratique linguistique ». Mais qu'on reconnaisse ou non ce caractère « normatif » de la signification, toute théorie systématique doit pouvoir rendre compte du fait de la compétence sémantique, telle qu'il est résumé par les thèses (2), (3) et (4) du philosophe imaginé par Schiffer. J'ai tenté de montrer qu'elle ne pouvait le faire qu'en prenant en compte certains faits psychologiques et cognitifs, sans que cela implique que, comme le voudrait la thèse (8) de Schiffer, la connaissance des significations se réduise à de tels faits.

Nous avons, enfin, rencontré un troisième type de difficultés, relatives à un autre aspect de l'opposition entre une forme de réalisme et une forme d'antiréalisme. Elles portent sur les thèses (1) et (5) de Schiffer, c'est-à-dire sur la réalité et l'objectivité de la signification et des contenus intentionnels : les faits de signification et les faits intentionnels existent-ils ? Bien que la position davidsonienne puisse être considérée comme un réalisme au sens dummettien, ses critiques la considèrent en fait plutôt comme un antiréalisme quant à la nature de la signification et du mental. Selon Davidson, la signification et les contenus mentaux sont ce qui se révèle à travers une procédure d'interprétation radicale, et sont donc relatifs à nos attributions. Ils semblent donc ne pas avoir d'existence objective en dehors de cette procédure et de ces attributions. Et puisque l'interprétation radicale est nécessairement indéterminée et holistique, il semble qu'on ne puisse jamais pouvoir dire quelle est *la* signification d'une expression ou *le* contenu sémantique d'une croyance. C'est, comme on l'a vu, la conséquence que n'hésitent pas à tirer les critiques du holisme de Davidson. Davidson est-il un antiréaliste, un instrumentaliste, voire un éliminativiste quant aux faits de significations et aux faits intentionnels en ce sens ? Non. Il soutient au contraire que les conditions de l'interpréta-

tion doivent réduire l'indétermination, et qu'il existe un accord intersubjectif possible à la fois sur la signification et sur la réalité. Il rejette l'idée que parce que le contenu sémantique ou psychologique est nécessairement interprété il n'y a pas objectivement de contenu sémantique ou psychologique. L'interprétation est un instrument de « mesure du mental », mais ce n'est pas parce qu'il y a plusieurs manières de mesurer qu'il n'y a rien à mesurer. En ce sens, Davidson n'est pas un sceptique ou un irréaliste quant aux contenus.

Mais le point de vue interprétatif sur la signification et les contenus mentaux ne s'identifie pas pour autant avec un réalisme intentionnel. Si Davidson a raison, le réalisme intentionnel est aussi injustifié, pour les contenus sémantiques et psychologiques, que le réalisme « externe » en métaphysique. Parce que ces contenus doivent toujours être soumis à des conditions normatives et holistiques, ils doivent être toujours *moins* objectifs que ne le soutient le réalisme sémantique ou le réalisme intentionnel, ils ne peuvent se prêter à une « naturalisation » ou à une réduction. On doit donc rejeter l'une des inférences de Schiffer (et de Kripke) : celle qui consiste à inférer, de la fausseté de la thèse réductionniste (8), que (1), le réalisme sémantique, et (5), le réalisme intentionnel, sont faux également, ou ne sont vrais que parce qu'il faut bien conserver nos croyances communes à ce sujet. Mais on doit aussi admettre que (1) et (5) ne sont vraies qu'en un sens minimal. Si le réalisme minimal défendu ici est correct, il faut dire, avec Wright (1989 : 247), que la signification et les contenus mentaux sont « dépendants de nos jugements et de nos réponses » mais que « bien que nos jugements [sur la signification et les contenus mentaux] n'aient pas d'épistémologie substantielle [ne soient pas totalement justifiables objectivement], ils peuvent néammoins être encore *objectifs*... ».

Cela reste compatible avec une forme de position interprétative. Le point où la position interprétative fait problème est le suivant. Elle réduit les contenus mentaux à des contenus d'attitudes propositionnelles et les tient comme intrinsèquement liés au langage et aux énonciations linguistiques. Cela lui interdit de prendre un point de vue explicatif authentique sur ces contenus, et de prendre en compte *d'autres* formes de contenus mentaux que ceux des croyances et des jugements. Or il n'y a aucune raison

de supposer qu'une théorie de l'interprétation doive se limiter à ces contenus. Si les considérations avancées ici au sujet de la connaissance tacite sont correctes, une théorie de la compétence sémantique doit recourir à d'autres formes de contenus. Il s'ensuit qu'on ne peut maintenir la thèse d'une interdépendance stricte de la pensée par rapport au langage, et qu'on doit rejeter, au moins sur ce point, ce que Dummett appelait « l'article de base de la philosophie du langage », selon lequel la philosophie du langage est le seul mode d'accès possible à la nature des pensées.

La question difficile, si l'on suit cette voie, devient celle de savoir comment on peut analyser la nature de ces contenus qui ne sont *pas* intentionnels ou sémantiques, alors même que *notre* notion usuelle de contenu ou de signification a des caractéristiques (holistiques, normatives) qui ne peuvent pas s'appliquer à ces « contenus ». Davidson a raison : la limite à laquelle se heurtent le réalisme intentionnel et une perspective « naturaliste » sur les contenus sémantiques et psychologiques est leur caractère normatif. Mais il admet aussi que ces contenus « surviennent sur », ou dépendent, de traits naturels. Le problème est qu'on ne peut pas se contenter d'affirmer cette survenance. Il faut encore montrer en quoi elle consiste, c'est-à-dire montrer comment la signification peut être à la fois un phénomène naturel et normatif. A cette question, je n'ai pas répondu, mais j'espère néanmoins avoir donné une mesure de la difficulté.

Références bibliographiques

Appiah A., 1986, *For Truth in Semantics*, Oxford, Blackwell.
Ayer A. J., 1936, *Language, Truth and Logic*, London, Golanz.
Baker G. & Hacker P., 1984, *Language, Sense and Nonsense*, Oxford, Blackwell.
— 1984a, *Scepticism, Rules and Language*, Oxford, Blackwell.
Baldwin T., 1982, « Prior and Davidson on Indirect Speech », *Philosophical Studies*, 42, 255-282.
— 1991, « Can there be a Substantive Account of Truth ? », *in* Engel & Cooper 1991d, 21-39.
Barrett R.B. et Gibson R.F., ed., 1990, *Perspectives on Quine*, Oxford, Blackwell.
Barwise J. et Perry J., 1983, *Situations and Attitudes*, Cambridge, Mass, MIT Press.
Bennett J., 1976, *Linguistic Behaviour*, Cambridge, Cambridge University Press.
— 1985, « Critical Notice of Davidson 1984 », *Mind*, XCVI, 549-570.
— 1988, *Events and their Names*, Cambridge, Cambridge University Press.
Bilgrami A., 1986, « Meaning, Holism and Use », *in* Le Pore, 1986, 101-124.
— 1992, *Belief and Meaning*, Oxford, Blackwell.
— 1992a, « Why Holism is Harmless and Necessary » tr. fr., à paraître *in* Engel, 1994.
Blackburn S., 1984, *Spreading the Word*, Oxford, Oxford University Press.
— 1985, « The Individual Strikes Back », *Synthese*, 58, 281-301.
— 1987, « Realists and Anti-realists », *Times Literary Supplement*, 27 feb.
— 1989 « Manifesting Realism », Midwest Studies in Philosophy, *14*, University of Notre Dame Press, 29-47.
— 1993, *Essays on Quasi-realism*, Oxford University Press.
Boghossian P., 1989, « The Rule Following Considerations », *Mind*, 507-549.
— 1990, « The Status of Content », *Philosophical Review*, XCIX, 157-183.
Bouveresse J. & Parret H., dir., 1981, *Meaning and Understanding*, Berlin, de Gruyter.

Bouveresse J., 1981, « Herméneutique et linguistique », *in* J. Bouveresse et H. Parret, 1981, repris dans Bouveresse, *Herméneutique et linguistique*, l'Eclat, 1991.
— 1986, « Le paradoxe de la règle », *Sud*.
— 1990, *Le pays des possibles*, Paris, Minuit.
— 1992, « La "causalité" des raisons », in *La raison, proche et lointaine, Mélanges en l'honneur de J.-P. Leyvraz, Studia Philosophica* 51, 33-60.
Burge T., 1979, « Individualism and the mental », *Midwest Studies in Philosophy*, 4.
— 1986, « Davidson on Saying that », *in* Le Pore, 1986.
Brandl J. et Gombocz W. L., eds, 1989, *The Mind of Donald Davidson*, Amsterdam, Rodopi.
Campbell J., 1986, « Conceptual Structure », *in* Travis, 1986, 159-174.
Canto M., 1991, dir., *Les paradoxes de la connaissance*, Paris, O. Jacob.
Carnap R., 1956, *Meaning and Necessity*, Chicago University Press.
Child T. W., 1987, « Critical Notice on Le Pore » 1986, *Mind*, XCVI, 549-570.
Chomsky N., 1986, *Knowledge of Language*, Praeger.
Church A., 1950, « On Carnap's Analysis of Statements of Assertion and Belief », *Analysis*, 10, 97-99.
Clementz F., 1985, « Sémantique formelle et philosophie des sciences », *Philosophie*.
Couture J., 1991, « Le molécularisme : logique et sémantique », in Laurier (dir.) 1991b, 159-180.
Cresswell M. C., 1976, « Semantic Competence », *in* C. Guenthner et C. Rohrer, eds, *Meaning and translation*, London, Duckworth.
— 1986, *Adverbial modification*, Dordrecht, Reidel.
Davidson D., 1957, *Decision making, an experimental Approach* (avec P. Suppes et S. Siegel), Stanford University Press (reprint : Chicago University Press, 1977).
— 1963 « Actions, Reasons, and Causes », *Journal of Philosophy*, 60, 685-700 (Davidson, 1980).
— 1966, « Theories of Meaning and Learnable Languages », *Proceedings of the 1964 International congress of Logic, Methodology and Philosophy of Science*, North Holland, 1966, 383-394 (Davidson, 1984).
— 1967, « Truth and Meaning », *Synthese*, 17, 304-323 (Davidson, 1984).
— 1967a « The Logical Form of Action Sentences », *The Logic of Decision and Action*, ed., N. Rescher, University of Pittsburgh Press, 81-95 (Davidson, 1980).
— 1967c « Causal Relations », *Journal of Philosophy*, 64, 691-703 (Davidson, 1980).
— 1968, « On Saying that », *Synthèse*, 19, 130-146 (Davidson, 1984).
— 1969, « True to the Facts », *Journal of Philosophy*, 66, 748-764 (Davidson, 1984).
— 1969a, « The Individuation of Events », *Essays in Honour of C. Hempel*, Rescher, N. ed., 216-234 (Davidson, 1980).
— 1969b, « How is Weakness of the will Possible ? », in *Moral Concepts*, ed., J. Feinberg, 93-113 (Davidson, 1980).

— 1970, « Semantics for Natural Languages », *Languaggi nell societa e nella Technica*, Communita, Milan, 177-188 (Davidson, 1984).
— 1970a « Mental Events », *Experience and Theory*, ed., L. Foster et J. Swanson, University of Massachusetts Press, Amherst, 79-101 (Davidson, 1980).
— 1973, « Radical Interpretation », *Dialectica*, 27, 313-328 (Davidson, 1984).
— 1973a, « In Defense of Convention T », *in* Davidson, 1984.
— 1974, « On the Very Idea of a Conceptual Scheme », *Proceedings and Adresses of the American Philosophical Association*, 47, 5-20 (Davidson, 1984).
— 1974a, « Belief and the Basis of Meaning », *Synthese*, 27, 309-323 (Davidson, 1984).
— 1975, « Thought and Talk », *in* Guttenplan, 1975, 7-23 (Davidson, 1984).
— 1976, « Reply to Foster », *in* Evans & McDowell, 1976, 33-41 (Davidson, 1984).
— 1976a, « Hempel on Explaining Action », *Erkenntnis*, 10, 239-253 (Davidson, 1980).
— 1977, « Reality without Reference », *Dalectica*, 31, 247-258 (Davidson, 1984).
— 1977a, « The Method of Truth in Metaphysics », *Midwest Studies in Philosophy*, ed., P. French, T. Uehling & H. Wettstein, 2, 244-254 (Davidson, 1984).
— 1978, « What Metaphors Mean », *Critical Inquiry*, 5, 31-47 (Davidson, 1984).
— 1979, « The Inscrutability of Reference », *The Southwestern Journal of Philosophy*, 10, 7-19 (Davidson, 1984).
— 1979a, « Quotation », *Theory and Decision*, 11 : 27-40 (Davidson, 1984).
— 1979b, « Moods and Performances », *Meaning and Use*, ed., A. Margalit, D. Reidel (Davidson, 1984).
— 1980, *Essays on Actions and Events*, Oxford University Press, tr. fr. et préface de P. Engel, *Actions et événements*, Paris, PUF, 1993.
— 1980a, « Toward a Unified theory of Meaning and Action », *Grazer Philosophische Studien*, 2, 1-12.
— 1982, « Rational Animals », *Dialectica*, 36, 317-327, tr. fr. *in* Davidson 1991d.
— 1982a, « Paradoxes of Irrationality », *in* R. Wollheim & J. Hopkins, eds, *Freud, A collection of Critical Essays*, Cambridge University Press, tr. fr. *in* 1991d.
— 1982b, « Empirical Contents », *in* R. Haller, ed., *Schlick und Neurath*, Rodopi, Graz : 471-489.
— 1983, « A Coherence Theory of Truth and Knowledge » *Kant Oder Hegel*, ed., D. Henrich, Klett-Cotta, 423-438, cité d'après Lepore, E., dir. (1986), 307-319.
— 1984, *Inquiries into Truth and Interpretation*, Oxford University Press, tr. fr. ; P. Engel, *Enquêtes sur la vérité et l'interprétation*, Nîmes, Jacqueline Chambon, 1993.
— 1984a, « Communication and Convention », *Synthese*, 59, 3-17.
— 1984b, « First-Person Authority », *Dialectica*, 38, 101-110.
— 1985, « Replies to Essays », *in* Vermazen B. & Hintikka M., 1985, 195-254.
— 1985a, « Reply to Quine on Events », *in* Le Pore 1985, 172-176.
— 1985b, « A New Basis for Decision Theory », *Theory and Decision* 18, 87-98.
— 1986, « A Nice Derangement of Epitaphs », *Philosophical Grounds of Rationality*, ed., R. Warner et R. Grandy, 156-174, cité d'après Le Pore, 1986.

— 1987, « Knowing One's Own Mind », *Proceedings and Addresses of the APA*, 60, 441-458.
— 1987a, « Problems in the Explanation of Actions », *in* Pettit, P. R. et S. J. Norman, dir. (1987), 35-49.
— 1989, « What is Present to the Mind ? », Brandl J. et W. L. Gombocz, ed., 1989, 3-18, tr. fr. P. Sauret, « Qu'est-ce qui est présent à l'esprit ? » *Lieux et transformations de la philosophie, Les Cahiers de Paris VIII*, Paris, Presses Universitaires de Vincennes, 1991.
— 1989a, « The Conditions of Thought », Brandl J. et W. L. Gombocz, dir. (1989), 193-200.
— 1990, « The Structure and Content of Truth » (*Dewey Lectures*), *Journal of Philosophy*, 87, 279-328.
— 1990a, « Representation and Interpretation », *in* Said, K. *et al.*, ed. (1990), 13-26.
— 1990c, « Meaning, Truth, and Evidence », *in* Barrett R. B. et R. F. Gibson, ed., (1990), 68-7.
— 1990d, « Turing's Test », Said K. A. M. *et al.*, dir. (1990), 1-12.
— 1991, « Reply to Burge », *Journal of Philosophy*, 88, 664-666.
— 1991, « The Myth of the Subjective », Krausz M., dir. (1989), 159-17.
— 1991a, « Epistemology Externalized », *Dialectica*, 45, 2-3, 191-202.
— 1991b, « Three Varieties of Knowledge », *in* Griffiths A. P., dir. (1991), 153-166.
— 1991d, *Paradoxes de l'irrationalité*, tr. fr. introduction par P. Engel de Davidson, 1982, 1982a, et 1985c, Combas, L'Eclat.
— 1992, « Thinking Causes », *in* J. Heil et A. Mele, eds, *Mental Causation*, Oxford, Oxford University Press.
— 1992a, « En quel sens un langage est-il social ? », *in* A. Soulez et J. Sébestik, dir., *Wittgenstein et la philosophie d'aujourd'hui*, Paris, Klincksieck.
Davidson D. & Harman G., 1972, eds, *Semantics for Natural Languages*, Dordrecht, Reidel.
— 1975, eds, *The Logic of Grammar*, Dickenson, Encino Ca.
Davies M., 1981, *Meaning, Quantification and Necessity*, Londres, Routledge.
— 1986, « Tacit Knowledge and the Structure of Thought and Language », *in* Travis, 1986, 127-156.
— 1986a, « Cognition, Consciousness and Concepts », *inédit*.
— 1987, « Tacit Knowedge and Psychological Explanation : Can a five per Cent difference Matter ? », *Mind*, 96, 441-462.
— 1988, « La connaissance tacite : modularité et subdoxasticité », *in* Engel, 1988b.
— 1991, « Adverbial Modification and Perception Reports », *in* Cooper & Engel, 1991d.
Dennett D., 1987, *The Intentional Stance*, Cambridge, Mass., MIT Press, tr. fr. P. Engel, *La stratégie de l'interprète*, Paris, Gallimard, 1990.

— 1993, « Back from the Drawing Board » *in* Dahlbom, B. *Dennett and his Critics*, Oxford, Blackwell, 203-235.
Devitt M., 1984, *Realism and Truth*, Princeton, Princeton University Press.
Dretske F., 1988, *Explaining Behaviour*, Cambridge Mass MIT Press.
Dummett M., 1959, « Truth », *Proceedings of the Aristotelian Society*, LIX, 141-162, repris *in* Dummett, 1978, tr. fr. F. Pataut, in *Philosophie de la logique*, Paris, Minuit, 1992.
— 1973, *Frege, Philosophy of Language*, Londres Duckworth, (cité dans la 1re éd., 2e éd., 1978).
— 1975, « The Justification of Deduction », *Proceedings of the British Academy*, LIX : 3-34, repris *in* Dummett, 1978.
— 1975a, « What is a Theory of Meaning ? (I) », *in* Guttenplan, 1975 : 97-138.
— 1976, « What is a Theory of Meaning ? (II) », *in* Evans & McDowell, 1976 : 67-137.
— 1978, *Truth and Other Enigmas*, London, Duckworth.
— 1982, *The Interpretation of Frege's Philosophy*, London, Duckworth.
— 1982a, « Realism », *Synthese*, 52, 55-112.
— 1985, « What do I Know when I Know a Language ? », inédit, Stockholm.
— 1986, « The Philosophy of Thought and the Philosophy of Language », *in* Vuillemin, dir., *Mérites et Limites des méthodes logiques en philosophie*, Paris, Vrin.
— 1986a, « Comments on Davidson and Hacking », *in* Le Pore, 1986 : 459-476.
— 1987, « Replies », *in* Taylor, 1987, 219-330.
— 1990, *The logical Basis of Metaphysics*, Harvard, Harvard University Press (The William James Lectures).
Donnellan K., 1966, « Reference and Definite Descriptions », *Philosophical Review*, 75, 284-304.
Engel P., 1981, « Davidson en perspective », *Critique*, 409-410, 578-594.
— 1982, « Quelques remarques sur la logique des phrases d'action » (avec F. Nef), *Logique et analyse*, 99, 291-319.
— 1985, *Identité et référence*, Paris, Presses de l'Ecole normale supérieure ; 1985a, « Comprendre un langage et suivre une règle », *Philosophie*, 1985, 8, 45-64.
— 1986, « Rapport sur « The philosophy of thought and the philosophy of language » du Pr Michael Dummett », *in* J. Vuillemin, dir., *Mérites et limites des méthodes logiques en philosophie*, Paris, Vrin, 155-163.
— 1986a, « Structure sémantique et forme logique d'après l'analyse aristotélicienne des phrases d'action », H. Joly, dir., *Philosophie et Grammaire dans l'Antiquité*, Grenoble et Bruxelles, *Recherches sur la philosophie et le langage* vol. 6-7 *et Revue de philosophie ancienne*.
— 1988, « Le sens littéral des métaphores », *Recherches sur la philosophie et le langage*, 9, 150-172.
— 1988a, « Radical Interpretation and the Structure of Thought », *Proceedings of the Aristotelian Society*, LXXXVIII, 1988, 106-177.

— 1988b, ed., *Psychologie ordinaire et Sciences cognitives*, Hermès III, Paris, CNRS.
— 1989, *La norme du vrai, philosophie de la logique*, Paris, Gallimard, 1989, tr. anglaise révisée par P. Engel et M. Kochan, *The Norm of Truth, An Introduction to the Philosophy of logic*, Hemel Hempstead, Harvester Wheatsheaf, 1991.
— 1989a, « Interpretation, charité et mentalité prélogique », *Revue philosophique*, 4, 543-558.
1990, « La philosophie du langage entre le clair et l'obscur », *Recherches sur la philosophie et le langage*, Hommage à Henri Joly, 12, 197-212.
— 1991, « Interpretation without Hermeneutics », *Topoi*, 10, 137-146.
— 1991a, « Adverbes, événements et structure sémantique », *Raisons pratiques*, 2, 1991, 229-250.
— 1991b, « La sémantique et le temps », Etude critique de F. Nef, *Sémantique de la référence temporelle*, L'âge de la science, IV, 1991.
— 1991c, « Holisme, molécularité et constantes logiques », *in* Laurier, 1991, 133-157.
— 1991d, ed. (avec N. Cooper), *New Inquiries into Meaning and Truth*, Hemel Hempstead, Harvester Wheatsheaf.
— 1992, *Etats d'esprit, questions de philosophie de l'esprit*, Aix-en-Provence, Alinéa 2e ed. augmentée *Introduction à la philosophie de l'esprit*, Paris, La Découverte, 1994.
— 1992a, « Rôle conceptuel et conditions de vérité », in *Langage et intentionalité*, dir., F. Lepage et D. Laurier, Montréal, Bellarmin-Vrin, 153-170.
— 1992b, « Actions, raisons et causes mentales », in *Philosophie de l'action*, dir., R. Glauser, *Revue de théologie et de philosophie*, 124, 3, 305-321.
— 1993, « L'antiréalisme réalisé », Etude critique de C. Wright, *Realism, Meaning and Truth*, L'Age de la science, 5, 77-90.
(à paraître a), « Davidson on Interpretation and Rationality », *in* Lewis Hahn, ed., *The Philosophy of Donald Davidson*, The Library of Living Philosophers, La Salle, Ill.
(à paraître b), « Intentionalité, interprétation et téléologie », *Actes du colloque sur l'intentionnalité*, dir., D. Janicaud, Nice.
(à paraître c) (ed.), *Lire Davidson*, Nîmes, *Editions de l'Eclat*.
(à paraître d), « Who's afraid of Meaning Holism ? »
Etchemendy J., 1988, « Tarski on Truth and Logical consequence », *The Journal of Symbolic Logic*, LIII, 51-79.
Evans G. & Mc Dowell J., eds, 1976, *Truth and Meaning, Essays in Semantics*, Oxford, Oxford University Press.
Evans G., 1976, « Semantic Structure and Logical Form », *in* Evans & Mc Dowell, 1976, 199-222.
— 1981, « Understanding Demonstratives », *in* Bouveresse & Parret, 1981 (Evans, 1985).
— 1981a, « Semantic Theory and Tacit Knowledge », *in* Holtzmann & Leich, 1981, (cité d'après Evans, 1985).
— 1982, *The Varieties of Reference*, Oxford, Oxford University Press.

— 1985, *Collected Papers*, Oxford, Oxford University Press.
Evnine S., 1991, *Donald Davidson*, Cambridge, Polity Press.
Field H., 1972, « Tarski's Theory of Truth », *Journal of Philosophy*, LXIX, 13, 347-375.
— 1974, « Quine and the Correspondence Theory », *Philosophical Review*, 83, 200-228.
— 1987, « The Deflationary Conception of Truth », *in* C. Wright et G. McDonald, eds, *Fact, Science and morality*, Oxford, Blackwell, 55-117.
— 1981, *Science Without Numbers*, Oxford, Blackwell.
— 1989, *Realism, Mathematics and Modality*, Oxford, Blackwell.
Fodor J., 1975, *The Language of Thought*, Cambridge, Mass., MIT Press.
— 1983, *The Modularity of Mind*, MIT Press, tr. fr. A. Gershenfeld, *La modularité de l'esprit*, Paris, Minuit, 1986.
— 1987, *Psychosemantics*, MIT Press, Cambridge University Press.
— 1990, *A Theory of Content and Other Essays*, Cambridge Mass, MIT Press.
Fodor J. & Pylyshyn Z., 1988, « Connexionism and Cognitive Architecture », *Cognition*.
Fodor J. A. et Lepore E., 1992, *Holism, A Shopper's Guide*, Oxford, Blackwell.
Follesdal D., 1979, « Hermeneutics and the Hypothetico-deductive method », *Dialectica*, 319-314.
Forbes G., 1987, « Truth, Correspondence and Redundancy », *in* C. Wright et G. Mc Donald, *Fact, Science and Morality*, Oxford, Blackwell.
Foster J., 1976, « Meaning and Truth theory », *in* Evans & Mc Dowell, 1976, 1-32.
Frege G., 1883, *Grundlagen der Arithmetik*, Breslau, tr. fr. C. Imbert, *Les fondements de l'arithmétique*, Paris, Seuil, 1969.
— 1893, *Grundgetzetze der Arithmetik*, Iéna, Réed. Olms, 1966, vol. 1.
George, A., 1990, « Whose Language is it anyway ? » *Philosophical Quarterly*, vol. 40, 160, 275-298.
Goldman A., 1979, « Interpretation Psychologized », *Mind and Language*, 4, 3, 161-185.
Goldfarb W., 1985, « Kripke on Wittgenstein and Rules », *Journal of Philosophy*, 82, 471-488.
Grandy R., 1973, « Reference, Meaning and Belief » *Journal of Philosophy*, 70, 439-452.
Grice H. P., 1957, « Meaning », *Philosophical Review*, 66, 377-388 (Grice, 1990).
— 1968, « Utterer's Meaning, Sentence-Meaning, and Word-Meaning », *Foundations of Language*, 4, 225-242 (Grice, 1990).
— 1969, « Utterers'Meaning and Intentions », *Philosophical Review*, 78, 147-177 (Grice, 1990).
— 1975 « Logic and Conversation », *in* Davidson & Harman, 1975 (Grice, 1990).
— 1990, *Studies in the Ways of Words*, Harvard, Harvard University Press.
Griffiths A. Phillips, dir., 1991, *A. J. Ayer: Memorial Essays*, Cambridge, Cambridge University Press.
Guttenplan S., 1975, ed., *Mind and Language*, Oxford, Blackwell.

Haack S., 1993, *Evidence and Inquiry, Towards Reconstruction in Epistemology*, Oxford, Blackwell.
Hacking I., 1986, « The Parody of Conversation », *in* Le Pore, 1986, 447-458.
Krausz M., dir., 1989, *Relativism*, Notre-Dame, University of Notre-Dame Press.
Harman G., 1975, « Meaning and Semantics », *in* M. Munitz, ed., *Logic and Ontology*.
— 1986, *Change in View*, MIT Press.
Heal J., 1986, *Fact and Meaning*, Oxford, Blackwell.
Higginbotham J., 1986, « Davidson's Program in Semantics », *in* Le Pore, 1986, 29-37.
Holtzmann and Leich, eds, 1981, *Wittgenstein, to Follow a Rule*, London, Routledge.
Horwich P., 1982, « Three Forms of Realism », *Synthese*, 51, 2, 181-201, 1990, *Truth*, Oxford, Blackwell.
Jacob P., 1992 « Le problème de l'esprit et du corps aujourd'hui », in D. Andler, ed. *Introduction aux sciences cognitives*, Paris, Gallimard, 1991.
Jeffrey R., 1965, *The Logic of Decision*, Chicago University Press, 2e éd., 1980.
Johnston M., 1988, « The End of the Theory of Meaning », *Mind and Language*, 3, 1, 28-42.
Kaplan D., 1977, « Demonstratives », *in* Almog *et al.*, *Themes from Kaplan*, Oxford University Press.
Katz J. et Postal P., 1964, *An Integrated Theory of Linguistic Descriptions*, Cambridge Mass, MIT Press.
Kenny A., 1963, *Action, Emotion and Will*, Oxford, Blackwell.
Kremer M., 1988, « Logic and Meaning the Significance of the Sequent Calculus », *Mind* 97, 51-79.
Kripke S., 1981, *Wittgenstein on Rules and Private Language*, Oxford, Blackwell, tr. fr. à paraître.
Laurier D., 1983, « Tarski, Davidson et la signification », *Dialogue*.
— 1985, « La langue d'une population : le lien entre sémantique et pragmatique », *Dialectica*.
Laurier Daniel, 1991a, « Comprendre ou interpréter ? », Laurier, D., dir. (1991b), 101-131.
Laurier D., dir., 1991b, *Essais sur le sens et la réalité*, Montréal/Paris, Bellarmin/Vrin.
Laurier D., 1994, « Holismes », à paraître *in* Engel, 1994.
Laurier D., 1994, *Introduction à la philosophie du langage*, Bruxelles, Mardaga.
Lepore Ernest et Brian McLaughlin, eds, 1985, *Actions and Events*, Oxford, Blackwell.
Lepore Ernest, ed., 1986, *Truth and Interpretation*, Oxford, Blackwell.
Le Pore E., 1982, « Truth and Inference », *Erkenntnis*, 18.
— 1983, « What Model-Theoretic Semantics cannot do », *Syntheses*, 54.
— 1984, « In Defense of Davidson », *Linguistics and Philosophy*, 5.
Le Pore E. et Loewer, B., 1989, « What Davidson should have Said », *in* Brandl & Gombocz, 1990, 65-78.
Le Pore E. et Loewer B., 1990, « You Can Say that Again », *in* Midwest Studies in Philosophy, 14.

Loar B., 1981, *Mind and Meaning*, Cambridge University Press.
Lycan W., 1984, *Logical Natural Language*, MIT Press, Cambridge, Mass.
Mackie J.-L., 1977, *Ethics, Inventing Right or Wrong*, Penguin.
Marr D., 1982, *Vision*, San Francisco, Freeman.
McDowell, 1976, « Truth-Conditions, Bivalence and Verificationism », *in* Evans & McDowell, 1976, 42-66.
— 1977, « On the Sense and the Reference of a Proper Name », *Mind*, 86, 159-185.
— 1981, « Anti-Realism and the Epistemology of Understanding », *in* Bouveresse & Parret, 1981, 225-248.
— 1981a, « Non Cognitivism and Rule-following », *in* Holtzman & Leich, 1981, 141-162.
— 1984, « Wittgenstein on Following a Rule », *Synthese*, 58, 325-363.
— 1985, « Functionalism and Anomalous Monism », *in* Le Pore, 1985, 387-398.
— 1987, « In Defense of Modesty », *in* Taylor, 1987, 59-80.
Mc Ginn C., 1977, « Charity, Interpretation and Belief », *Journal of Philosophy*, 74, 521-535.
— 1980, « Truth and Use » in Platts 1980.
— 1985, *Wittgensein on Meaning*, Oxford, Blackwell.
— 1986, « Radical Interpretation and Epistemology », *in* Le Pore, 1986, 356-368.
Millikan R., 1984, *Language, Thought, and Other biological Categories*, MIT Press, Cambridge Mass.
Montague R., 1974, *Formal Philosophy*, Yale, Yale University Press.
Nef F., 1986, *Logique et langage*, Paris, Hermès.
Nisbett R. et Thagard P., 1983, « Rationality and Charity », *Philosophy of Science*.
Panaccio C., 1992, *Les mots, les concepts et les choses, La sémantique de Guillaume d'Occam et le nominalisme d'aujourd'hui*, Paris-Montréal, Vrin-Bellarmin.
Pariente J.-C. et Charolles M., *La grammaire de Montague*, P. Lang, Berne.
Parsons T., 1972, « Some Problems concerning the Logic of Predicate Modifiers », *in* Davidson & Harman, 1972.
Peacocke C., 1976, « Truth Definitions and Actual Languages », *in* Evans & Mc Dowell, 1976, 162-188.
— 1981, *Holistic Explanation*, Oxford, Oxford University Press.
— 1981a, « The Theory of Meaning in Analytic Philosophy », *in* Floistad, ed., *Contemporary Philosophy*, vol. 1, La Haye, Nijhoff.
— 1983, *Sense and Content*, Oxford, Oxford University Press.
— 1986, *Thoughts, an Essay on Content*, Blackwell, Oxford.
— 1986a, « Explanation in Computational Psychology : Language, Perception and Level 1.5 », *Mind and Language*.
— 1988, « The Limits of Intelligibility : a Post-Verificationist Proposal », *Philosophical Review*, 97, 463-496.
— 1992, *A Study of Concepts*, Cambridge Mass, MIT Press.

Perry J., 1979, « Frege on Demonstratives », *Philosophical Review*, tr. fr. J. Dokic et E. Corazza, in *Penser en contexte*, Combas, l'Eclat, 1993.
Pettit P., 1990, « The Reality of Rule Following », *Mind*.
— 1991, « Realism and Response Dependence », *Mind*.
Pettit P. R. et S. J. Norman, dir., (1987), *Metaphysics and Morality*, Oxford, Blackwell.
Platts M., ed., 1980, *Reference, Truth and Reality*, London, Routledge.
— 1981, *Ways of Meaning*, Routledge, London.
Popper K. R., 1972, *Objective Knowledge*, Oxford, Oxford University Press, tr. fr. *La connaissance objective*, Ed. Complexe.
Prawitz D., 1977, « Meaning and Proofs : on the Conflict Between Classical and Intuitionistic Logic », *Theoria*, 43, 2-40.
Putnam H., 1975, « The Meaning of "Meaning" », in *Philosophical Papers II*, Cambridge University Press.
— 1978, *Meaning and the Moral Sciences*, London, Routledge.
— 1981, *Truth, Reason and history*, Cambridge University Press, tr. fr. A. Gerschenfeld, *Raison, Vérité, et Histoire*, Paris, Minuit, 1987.
— 1983, *Philosophical Papers, vol. 3*, Cambridge University Press ; (1983a).
— 1990, « On Truth », *in* L. S. Cauman *et al.*, eds, *How Many Questions*, Indianapolis, Hackett, 35-56.
— 1990, *Realism with a Human Face*, Harvard University Press.
Quine W. V. O., 1936, « Truth by Convention » in *The Ways of Paradox*, Harvard, Harvard *University Press*, 1976.
— 1940, *Mathematical Logic*, Cambridge, Mass., Harvard University Press.
— 1951, *From a Logical Point of View*, Harvard, Harvard University Press.
— 1960, *Word and Object*, Cambridge Mass., MIT Press, tr. fr. P. Gochet et J. Dopp, *Le mot et la chose*, Paris, Flammarion, 1977.
— 1970, *Philososophy of logic*, Prentice Hall, tr. fr. *Philosophie de la logique*, Paris, Aubier, 1976.
— 1972, « Methodological Reflections on Linguistic Theory », *in* Davidson & Harman, 1972.
Ramberg B., 1989, *Davidson's Philosophy of Language*, Oxford, Blackwell.
Ramsey F. P., 1926, « Truth and Probability », in *Foundations*, ed., H. Mellor, London, Routledge.
Rorty R., 1979, *Philosophy and the Mirror of Nature*, Princeton University Press, tr. fr. T. Marchaisse, *L'Homme spéculaire*, Paris, Seuil, 1990.
— 1986, « Pragmatism, Davidson and Truth », *in* Le Pore, 1985, tr. fr. J.-P. Cometti, *in* Rorty R., *Science et Solidarité*, Combas, L'Eclat.
Rosenberg A., 1985, « Davidson's Unintended Attack against Psychology », *in* Le Pore, 1985.
Said K. A. M. *et al.*, eds, 1990, *Modelling the Mind*, Oxford, Oxford University Press.

Sainsbury M., 1977, « Semantics by Proxy », *Analysis*.
Scheffler I., 1954, « An Inscriptional Approach to Indirect Quotation », *Analysis*, 10, 83-90.
Searle J., 1983, *Intentionality*, Cambridge, Cambridge University Press, tr. fr. C. Pichevin, *L'intentionalité*, Paris, Minuit, 1987.
— 1992, « Indeterminacy, Empiricism and the first Person », *Journal of Philosophy*, 84.
Schiffer S., 1987, *Remnants of Meaning*, Cambridge (Mass.), MIT Press.
Seymour S., 1991, « La sémantique de Davidson et le problème de la compréhension », *in* Laurier, 1991.
— 1992, « Revue critique de Engel, 1989 », *Dialogue*, 1992.
— 1994, « Discours indirect et citation », à paraître *in* Engel, 1994, et Seymour, 1994.
— 1994a, *Pensée, Langage et Communauté*, à paraître.
Skorupski J., 1988, « Review of Wright, 1987 », *Philosophical Quarterly*, 38 : 500-525 : 1992.
« Anti-Realism, Inference and the Logical Constants », in Haldane J. and Wright C., 1993, *Realism and Reason*, Blackwell, Oxford.
Smith B. C., 1992, « Understanding Language », *Proceedings of the Aristotelian Society*, XCII, 109-141.
Soames S., 1984, « What is a Theory of Truth ? », *Journal of Philosophy*, LXXXI, 8, 411-429.
— 1985, « Semantics and Psychology », *in* J. Katz, ed., *The Philosophy of Linguistics*, Oxford University Press.
Sperber D. et Wilson D., 1986, *Relevance*, Oxford, Blackwell, tr. fr. D. Sperber et A. Gerschenfeld, *La pertinence*, Paris, Minuit, 1990.
Stampe D., 1977, « Towards a Causal Theory of Linguistic Representation », *in* French & *alii*, 1977.
Stich S., 1976, « Davidson's Semantic Programme », *Canadian Journal of Philosophy*.
— 1978, « Beliefs and Subdoxastic States », *Philosophy of Science*.
— 1983, *From folk Psychology to Cognitive Science*, Cambridge, Mass., MIT Press.
Strawson P. F., 1970, « Meaning and Truth », in *Logico-linguistic Papers*, Methuen, tr. fr. *Papiers logico-linguistiques*, Paris, Seuil.
— 1976, « On Understanding the Structure of One's Language », *in* Evans & Mc Dowell, 1976.
Tarski A., 1956, *Logic, Semantics, Metamathematics*, Oxford, Oxford University Press, tr. fr. G. Granger et alii *Logique, Sémantique, métamathématique*, Paris, A. Colin, 1972.
Taylor B., 1985, *Modes of Occurrence*, Oxford, Blackwell, ed., 1987, *Michael Dummett, Contributions to Philosophy*, La Haye, Nijhoff.
Tennant N., 1977, « Truth, Meaning and Decidability », *Mind*, 86, 368-387.
— 1987, *Anti-realism and logic*, Oxford, Oxford University Press.
— 1987a, « Holism, Molecularіy, and Truth », *in* Taylor, 1987.

Tiercelin C., 1993, *La pensée-signe, études sur Peirce*, Nîmes, J. Chambon.
Van Fraassen B. C., 1980, *The Scientific Image*, Oxford, Oxford University Press.
Vermazen B. et Merrill B. Hintikka, dir., 1985, *Essays on Davidson: Actions and Events*, Oxford, Oxford University Press.
Wallace J., 1977, « Only in the Context of a Sentence Do Words Have any Meaning », *in* French, Uehling et Wettstein, 1977.
Weinstein S., 1974, « Truth and Demonstratives », *Noûs*, 8, 179-184, repris dans Davidson et Harman, 1975.
Wheeler S., 1986, « Indeterminacy of French Interpretation : Davidson and Derrida », *in* Le Pore, 1986, 477-494.
Wiggins D., 1976, « Truth, Invention, and the Meaning of Life », *in* Wiggins, 1986.
— 1980, « What would be a Substantial Theory of Truth ? », *in* Z. Van Staaten, ed., *Philosophical Subjects, Essays in honor of P. F. Strawson*, Oxford, Oxford University Press.
— 1986, *Needs, Values, Truth*, Oxford, Blackwell.
Williams M., 1986, « Do we (Epistemologists) Need a Theory of Truth ? », *Philosophical Topics*, XIV, 223-242.
Wilson N. L., 1959, « Substances without Substrata », *The Review of Metaphysics*, 12, 521-539.
Wittgenstein L., 1922, *Tractatus Logico-Philosophicus*, Routledge and Kegan Paul, tr. fr. G. Granger, Paris, Gallimard, 1993.
— 1953, *Philosophische Untersuchungen*, ed., Anscombe, Oxford, Blackwell.
— 1958, *Bemerkungen Uber die Grundlagen der Mathematik*, Oxford, Blackwell, tr. fr. M. A. Lescourret, *Remarques sur les fondements des mathématiques*, Paris, Gallimard, 1985.
Wright C., 1981, « Rule Following, Objectivity and the Theory of Meaning », *in* Holtzman & Leich, 1981.
— 1983, *Frege's theory of Numbers as Objects*, Aberdeen University Press.
— 1985, « Kripke's Account of the Argument Against Private Language », *Journal of Philosophy*, 81, 759-768.
— 1987, *Realism, Meaning and Truth*, Oxford, Blackwell, seconde éd. ; augmentée 1993 (références données d'après la première éd.).
— 1987a, « Anti-realism, Irrealism, Quasi-realism », *Midwest Sudies in Philosophy*, XII, 29-47.
— 1993, *Truth and Objectivity*, Harvard, Harvard University Press.

Index des noms

Index des notions

DU MÊME AUTEUR

Identité et référence, la théorie des noms propres chez Frege et Kripke, Paris, Presses de l'Ecole normale supérieure, 1985.

La norme du vrai, philosophie de la logique, Paris, Gallimard, 1989, traduction anglaise révisée par P. Engel et M. Kochan, *The Norm of Truth, an Introduction to the Philosophy of Logic*, Harvester Wheatsheaf, Hemel Hempstead, 1991.

Etats d'esprit, question de philosophie de l'esprit, Aix-en-Provence, Alinéa, 1992, 2e éd. augmentée, *Introduction à la philosophie de l'esprit*, Paris, La Découverte, 1994.

(Avec N. Cooper), éd., *New Inquiries into Meaning and Truth*, Harvester Wheatsheaf, Hemel Hempstead, 1991.

Lire Davidson (éd.), Combas, L'Eclat, 1994.