La définition

Pierre Wagner

La définition

Résumé

Les dialogues de Platon mettent en scène Socrate en quête de la définition générale d'une vertu (justice, courage, beauté, connaissance, etc.) auprès d'experts censés savoir « ce que c'est » mais qui se révèlent incapables de saisir l'essence de la chose en question. L'interrogation socratique soulève le problème de l'objectif de la définition et de sa place dans un discours qui vise la connaissance. La définition est en effet habituellement posée comme liminaire, principe qui éclaire l'objet dont il va être question. Or si elle prétend être davantage que le recueil lexicographique des significations fixées par l'usage pour énoncer ce que les choses sont en elles-mêmes, ne devrait-elle pas figurer au terme de la quête du savoir plutôt qu'au nombre de ses prémisses ?

Sans doute les définitions ont-elles des sens et des buts variés et de fait, les philosophes distinguent définitions nominales et réelles, descriptives et stipulatives, abréviatives et explicatives, implicites et explicites, analytiques et synthétiques, prédicatives et imprédicatives. Des règles de la définition sont également débattues qui, au-delà des exigences de clarté et d'univocité, garantissent les conditions de non créativité et d'éliminabilité du défini, caractéristiques de la théorie classique des définitions. Les exemples historiques ne manquent pas, en effet, d'essais de définitions qui, faute de respecter ces règles, se révélèrent vides d'objets ou engendrèrent la contradiction.

Les formes définitionnelles varient selon la nature du mot ou de la chose à définir et selon la visée de l'acte définitoire. Cela vaut non seulement dans le cas des langues « naturelles » mais également pour les langues formelles dont usent logiciens, mathématiciens et informaticiens. Appliquées aux définitions, les méthodes formelles ont pour vertu d'éclairer certaines questions classiques comme le problème des indéfinissables, les conditions de la définissabilité ou encore les vices et vertus des définitions circulaires.

NB : la lecture des paragraphes 9 à 13 est rendue plus aisée par une certaine familiarité avec les symboles en usage dans la logique élémentaire. Pour un examen plus approfondi des questions de définissabilité qui peuvent être soulevées dans un cadre formel, quelques-uns des textes auxquels le lecteur est renvoyé requièrent quant à eux une formation en logique plus avancée.

Table des matières

1. Limites de la définition

Au regard de l'étymologie, « définition » désigne l'acte de tracer ou de reconnaître une limite, ou le résultat de cet acte, qui circonscrit et sépare ce qui se trouve ainsi enclos. Le latin « *definitio* » est construit sur « *finis* » : la limite, la borne, la frontière. Au regard de l'usage, cet acte est verbal et vise à dire ce qui tombe, ou ne tombe pas, sous un concept : la délimitation s'opère par exposition d'une essence ou d'une signification. Définir, c'est énoncer la nature d'une chose ou le sens d'un terme. La définition du cercle dit ce qu'il est – un ensemble de points du plan géométrique, équidistants d'un même point appelé « centre » – et le sens des mots « détroquer », « centon » ou « Paraclet » se lit dans leur définition.

Pour autant, la définition n'est pas nécessairement requise pour l'appréhension ou l'explication d'une signification. D'autres méthodes sont à disposition, que de bonnes raisons font souvent préférer à la définition elle-même. Leurs traits distinctifs dessinent en creux le propre du définir. Elles ont recours aux figures archétypales, aux exemples, à l'image ou encore à l'usage. Découvrir dans le *Tartuffe* le paradigme de l'hypocrisie, apprendre à distinguer drame et tragédie par la fréquentation des salles de théâtre, observer les planches qu'une encyclopédie consacre au dodo, ou saisir en contexte, à l'occasion d'échanges verbaux et de lectures, les usages nuancés d'adjectifs comme « matois » ou « madré » concourent à l'intelligence de la langue sans offrir à proprement parler aucune définition des termes en question. Ou encore : une liste de synonymes qui avoisinent le sens d'un mot sans en atteindre l'exactitude et une série d'antonymes qui en esquissent le contre-pied ; procédés qui confèrent, dans la plupart des cas, toute la lumière qu'exigent les circonstances, bien qu'ils ne répondent pas davantage que les précédents à la demande d'une définition.

La définition est elle-même un concept aux frontières floues et aux usages variés en sorte qu'il n'est pas si aisé de dire en toute rigueur quels procédés en relèvent et quelles méthodes elle exclut. On peut hésiter, par exemple, à juger que l'attribution d'un nom propre ait valeur de définition. Reconnaîtra-t-on un mode particulier du définir dans le lien entre « Pont du Gard » et l'ouvrage architectural ainsi nommé, ou entre le nom « Aristote » et l'auteur de l'*Éthique à Nicomaque* ? Un nom propre est-il susceptible d'être *défini* ? La réponse pourrait dépendre du mode d'association du nom et de la chose, du fait qu'un mouvement de la main puisse accompagner la désignation de l'objet dans le premier cas alors qu'une description définie comme « le philosophe de Stagire » soit requise dans le second pour décider qui le nom nomme. Ce point fait en réalité débat car à en croire Kripke, même dans le cas d'un être absent et inaccessible à la désignation directe par un geste à valeur baptismale ou informative, une description singularisante peut tout à fait servir à *fixer la référence* du nom sans aucunement lui *conférer un sens* (Kripke 1980, 20 et 46). Or fixer ou dire quelle est la référence d'un nom propre sans qu'il devienne de ce fait porteur de sens, est-ce à proprement parler *définir* ?

On répondra sans doute qu'il s'agit alors d'une manière de définition *par ostension*, procédé applicable autant aux couleurs, aux goûts et autres *qualia* qu'à l'objet « Pont du Gard ». Ce qui est débattu, dans ce cas, est la possibilité même d'une telle définition de « rouge » ou de « sucré » si le mot n'est pas déjà connu par d'autres voies, car la pointe de l'index orienté vers l'objet peine à tracer les limites exactes de ce qui est désigné, par intention, et le complément discursif qui pourrait

expliciter la visée de sens ferait retomber, par définition, hors de la pure ostension. Est-il pour autant exclu, comme l'affirme le Philalèthe des *Nouveaux essais sur l'entendement humain*, de produire une authentique définition nominale de nos « idées simples », tout comme il semble exclu de « connaître le goût de l'ananas par la relation de nos voyageurs » (Leibniz 1765, livre III, ch. 4, § 11). Ces cas limites illustrent les frontières incertaines de la définition et mettent à l'épreuve toute tentative pour en saisir l'essence, tout effort pour en circonscrire précisément la signification.

2. Les apories du définir

Dans les cas moins équivoques, réduite à sa plus simple expression, la définition d'un certain *x*, quel qu'il soit, se caractérise généralement comme une identité ou une équivalence : entre *x* d'une part et, d'autre part, ce qu'il signifie ou l'essence qu'il exprime. Caractérisation dont l'apparente simplicité ouvre pourtant le champ de nouveaux questionnements et de toute une cohorte d'interrogations aporétiques. De quoi, pour commencer, y a-t-il définition ? Définit-on les choses ou les mots ? Et quelles conséquences une telle différence induit-elle ? Philosophes et logiciens se sont attachés à distinguer définitions de noms et de choses, « nominales » et « réelles », adjoignant parfois à cette catégorisation simple de nouvelles subdivisions. Si l'on admet que définir *x* consiste à en présenter la signification (que *x* soit mot ou chose), quelle relation existe-t-il entre le *definiendum* (littéralement : ce qui doit être défini) et le *definiens* (ce qui sert à définir ; littéralement : le « définissant ») ? Qu'il s'agisse d'une identité ou d'une équivalence, en quel sens faut-il entendre cette relation ? L'alternative renvoie à de nouvelles apories, sur la nature de l'identité, sur le sens de l'équivalence, ou sur la signification elle-même, dont il est notoirement difficile de dire au juste ce qu'elle est. Ou encore sur la *forme* de la définition, déjà en question dans le *Sophiste* de Platon ou dans les débats anciens sur les procédés de saisie d'une essence par division du genre en espèces et sous-espèces (Aristote 2005, livre II, chap. 13), question renouvelée au début du XX^{e} siècle par les discussions relatives aux définitions mathématiques qui visent des objets hautement abstraits ou des totalités infinies.

Autres difficultés : que vise la définition ? Toutes les définitions ont-elles un seul et même objectif ? La lexicographe et la naturaliste satisfont-elles la même interrogation lorsque l'une et l'autre prétendent donner par une définition la signification des mots « or » et « argent » ? Quant aux activités définitoires respectives du législateur, de la mathématicienne, du philosophe, de la physicienne, leurs buts sont-ils communs ? Une seule et même motivation anime-t-elle Euclide lorsqu'il pose les définitions liminaires du point, de la droite ou du cercle dans ses *Éléments* et Socrate dans sa quête d'une définition du juste, du beau ou du pieux ? La diversité des visées de la définition, selon qu'elle a valeur de stipulation, de description ou d'explication pourrait bien induire un éventail irréductible de formes définitionnelles variées.

De quoi, par ailleurs, les définitions dépendent-elles ? Du choix d'une langue partagée, au sein de laquelle les termes sont définis les uns à partir des autres au moyen de ressources lexicales, syntaxiques et sémantiques variables selon les idiomes ? D'un vocabulaire de base, d'abord soigneusement circonscrit, puis méthodiquement élargi par la définition de mots nouvellement introduits ? Ou encore d'une théorie d'arrière-plan qui détermine, au moins partiellement, le sens même des termes de base du langage retenu ? De là, également, le problème des limites : est-il

possible de tout définir ou existe-t-il, et en quel sens, des indéfinissables ? Le sens de nombreux termes semble également parfois trop vague, et leur usage trop fluctuant, pour être précisément circonscrits. La forme d'une définition pourrait aussi dépendre de la catégorie du signe à définir, car on imagine mal définir « quelque » ou « bien que » en usant des mêmes procédés que pour « tigre » ou « démocratie ». Comment définir, enfin, un terme qui n'a de signification qu'en contexte, tels les syncatégorématiques du logicien ou les explétifs du grammairien ? Certains philosophes ont étendu cette difficulté à tous les mots, tel Frege qui soutient, dans l'ouvrage qu'il consacre à la définition du nombre, que « les mots n'ont de signification qu'au sein d'une proposition ». S'agit-il donc, s'il en est ainsi, de définir un mot non pas isolément mais dans le contexte propositionnel où il figure ? (Frege 1969, Introduction et § 62).

Il arrive aussi qu'une formulation à visée définitionnelle échoue à définir quoi que ce soit. Ainsi « le plus petit entier naturel qui ne peut être défini en moins de vingt-sept syllabes », formulation qui a toutes les apparences d'un authentique *definiens* et qui se trouve comprendre exactement vingt-six syllabes. Si l'on supposait qu'elle caractérise effectivement un certain entier, il faudrait conclure que celui-ci ne satisfait pas la propriété censée le définir, à savoir d'être *indéfinissable en moins de vingt-sept syllabes*. Ce prétendu *definiens* ne définit donc rien. Semblable aporie, connue sous le nom de « paradoxe de Berry » (Russell 1906, 645), invite à réfléchir aux règles de la définition et aux conditions de la définissabilité, car selon toute vraisemblance, la difficulté réside dans l'expression « ne peut être défini » qui, en dépit de son apparente clarté, pèche par un singulier défaut de précision sur ce qui est susceptible de valoir comme une définition acceptable.

La chaîne des questions que soulève la définition, qui touchent au premier chef les langues usuelles dites « naturelles », est redoublée par leur application aux langues formelles dont informaticiens, logiciens et autres analystes font aujourd'hui communément usage. Les difficultés classiques de la définition se trouvent alors redoublées, car portées à l'horizon de méthodes formelles et formulées en des termes qui les ouvrent à une possible exactitude mathématique, qu'il s'agisse de la relation entre *definiens* et *definiendum*, des indéfinissables, des distinctions entre définition implicite et explicite, stipulative ou explicative, nominales ou réelles, prédicatives ou imprédicatives, des définitions circulaires, de l'éliminabilité des termes définis, ou des conditions générales d'une définition logiquement correcte. Le traitement logique de tels problèmes, adapté aux langues formelles, n'est généralement pas accessible à une transposition directe aux langues naturelles, dont les contours fluctuants et vagues rendent malaisée toute visée d'une solution exacte. Pour autant, les approches logiques de la définition – on en distingue plusieurs – éclairent dans bien des cas les questions traditionnelles précédemment évoquées en produisant certains modèles mathématiques ou formels qui servent utilement de point de comparaison et de référence aux réflexions sur la nature, le sens et les fonctions des définitions dans les langues naturelles.

Sur la question de la définissabilité comme sur d'autres, les méthodes formelles ont également vertu clarificatrice, y compris pour des questions que les philosophes ont soulevées dans un contexte éloigné de tout traitement mathématique. Il semble néanmoins préférable, avant de mobiliser les outils conceptuels de l'analyste, d'exposer quelques problèmes généraux de la définition et de préciser en quels termes ils ont pu être soulevés. Mais comme le suggèrent les remarques précédentes, le prix à payer est un renoncement à toute théorie unifiée. Un examen superficiel suffit à révéler la variété des formes et des buts de l'activité définitoire, remarque qui vaut tant pour les

techniques de définitions élaborées dans un cadre formel que pour les définitions données dans une langue naturelle, sans compter la définition des concepts (tels *vrai, possible, nécessaire* ou *identique*) qui semblent requérir une approche spécifique et un traitement séparé. Le rappel de quelques principes classiques de classification des définitions pourra au mieux, pour commencer, contribuer à diminuer, en l'éclairant, la diversité du champ des exemples que nous offre l'usage. Les approches formelles, souvent aptes à clarifier ou affiner l'analyse, n'en sont pas moins, elles aussi, essentiellement plurielles en sorte que leur examen mérite également de précautionneuses distinctions.

3. Définitions de nom, de mot et de chose

Une définition du plomb est susceptible de répondre à des interrogations distinctes selon qu'elle porte sur l'essence de la chose ou la signification du mot ; sur ce que le plomb *est* en lui-même ou sur ce qu'on *entend* par « plomb » à l'aune d'un certain usage. Dans le premier cas, on suppose qu'un échantillon du métal en question est donné, puis soumis à observation et analyse – optique, mécanique, chimique ou autre – dont l'objet est la matière même de l'échantillon, non le mot qui la désigne. La réponse dépend de la chose ainsi isolée et des méthodes d'examen appliquées, non du mode de désignation. Dans le second cas en revanche, la question, relative à un usage particulier, est susceptible de varier avec la langue à laquelle le mot « plomb » est emprunté et la communauté linguistique considérée. On ne s'étonnera donc pas que les définitions de la lexicographe et du physicien puissent ne point coïncider.

Une autre différence tient au fait que la définition d'une espèce naturelle ou d'un composé est aussi susceptible d'évoluer avec les capacités d'analyse dont disposent physiciens et chimistes, comme le montrent, en histoire des sciences, les cas de l'oxygène ou des acides. L'histoire des sciences abonde aussi en exemples de définitions qui se révèlent ultérieurement vides car sans objet. On pense à celles du phlogistique dans l'ancienne chimie, de l'éther des physiciens du XIX^{e} siècle, ou de la planète Vulcain, censée rendre compte de l'avance du périhélie de Mercure. On imagine mal pareille mésaventure pour une définition dont le seul objectif est de recueillir un usage au sein d'un groupe de locuteurs.

La distinction qu'on vient d'esquisser ne recoupe cependant pas l'opposition classique qu'on trouve par exemple chez Pascal ou dans la *Logique de Port-Royal*, entre définitions de nom et de chose. Le modèle de la définition de nom est en effet celle du géomètre qui procède par « impositions de nom aux choses qu'on a clairement désignées en termes parfaitement connus » ; par exemple « j'appelle tout nombre divisible en deux également, nombre pair » (Pascal 1954, 577). Elle ne procède donc ni d'une enquête relative à l'usage, ni de l'analyse d'un échantillon ou d'un exemplaire. Ainsi comprise et opposée à « définition de chose », l'expression « définition de nom » risque de prêter à confusion et requiert donc une attention particulière : il n'est en effet nullement exclu, bien au contraire, que son *definiendum* soit précisément une *chose*.

La logique de Port-Royal distingue la définition de nom de « l'explication de ce qu'un mot signifie selon l'usage ordinaire d'une langue » (Arnauld et Nicole 1992, partie I ch. 12), ce que les historiens de la science du langage nomment parfois « définition de mot » pour la distinguer de la définition de

nom (Auroux 1990, 32). Quant à la définition de chose, son principe remonte à l'idée d'Aristote selon laquelle la définition dit l'essentiel de ce que la chose définie est. Cette conception de la définition en fait un énoncé susceptible d'être vrai ou faux, et non une convention. Pour Pascal au contraire « les définitions ne sont faites que pour désigner les choses que l'on nomme, et non pas pour en montrer la nature » (Pascal 1954, 580). En conséquence, ce qu'il dit de la définition touche, pour l'essentiel, à la définition de nom.

L'expression « définition de chose » est du reste largement tombée en désuétude. Il arrive qu'on parle de définition *analytique* lorsque son objet est un composé ou une espèce naturelle, par opposition aux définitions *synthétiques* de concepts ou d'objets résultants de constructions mathématiques (Czeżowski 2000). Cela vaut non seulement pour des constructions qui sont clairement le fruit de l'inventivité des seuls mathématiciens, mais aussi pour la structure des entiers naturels – qui semble effectivement si naturelle que beaucoup la tiennent pour donnée dans l'intuition – ou pour la définition du nombre π comme rapport de la circonférence d'un cercle à son diamètre dans le plan euclidien. Savoir s'il s'agit là de définitions de noms ou de choses pourrait bien dépendre d'options ouvertes en philosophie des mathématiques, où réalistes et anti-réalistes s'opposent depuis l'Antiquité.

Ces distinctions sont à mettre en rapport avec la position de Kant qui affirme qu'« un concept empirique ne peut être défini, mais seulement explicité » et qui oppose par ailleurs définitions philosophiques et mathématiques. Alors qu'en philosophie les définitions consistent en « analyses de concepts donnés », qu'il n'est donc possible de produire qu'au terme de l'analyse, en mathématiques, « le concept est d'abord donné par la définition », en sorte que les définitions sont essentiellement liminaires et que « nous n'avons absolument aucun concept avant la définition » (Kant 1980, A730-731, B758-759).

La distinction entre définition « nominale » et définition « réelle » que nombre de philosophes et logiciens ont reprise, a des motivations et des caractérisations variables d'un auteur à l'autre, qui ne recouvrent du reste que partiellement celles qu'on vient d'esquisser. On s'en convaincra aisément en comparant par exemple ce que Guillaume d'Ockham dit dans sa *Somme de logique* de la « *definitio exprimens quid rei* » et de la « *definitio exprimens quid nominis* » (littéralement : définition exprimant ce qu'il en est de la chose et, respectivement, ce qu'il en est du nom) (Ockham 1993, chap. 10 et 26), aux différences établies par Jean Buridan entre quatre modes de la définition selon qu'elle est nominale, quidditative, causale ou descriptive (Buridan 2001, 631-663), ou encore aux textes dans lesquels Leibniz distingue définition réelle et définition nominale, la première montrant la possibilité de la chose définie tandis que la seconde n'est qu'énumération de caractères suffisant à sa reconnaissance (Leibniz 1684 ; 1686 § 24 ; 1765, livre I ch. 12).

Aristote reconnaît lui aussi différentes sortes de définitions mais il n'est pas du tout évident que celle qu'il qualifie de « nominale » (« onomatōdes ») dans les *Seconds analytiques* (Aristote 2005, livre II, chap. 10, 93b31) soit apparentée à celles que la tradition philosophique ultérieure nomme ainsi car il n'y a, pour Aristote, de définition que de ce qui est. On peut bien dire ce que signifie le mot « bouc-cerf » mais on n'aura pas pour autant, aux yeux du Stagirite, donné une définition ; car comment énoncer l'essence, le « ce que c'est », de ce qui n'est pas ?

4. Définitions stipulatives, descriptives, explicatives

Une définition peut avoir pour objet de recueillir les sens d'un terme dont l'usage ne résulte d'aucune décision, ou du moins d'aucune décision connue. Lorsque le travail de définition est relatif à une communauté linguistique dont la langue, telle une donnée objective, est soumise à observation, les énoncés qui en résultent ont valeur *descriptive* et se distinguent alors des définitions *stipulatives*. Les linguistes quant à eux affinent encore ces différences et distinguent définitions terminologiques, lexicographiques et encyclopédiques (de Bessé 1990).

S'il n'est certes pas exclu que les définitions stipulatives se conforment à quelque usage, elles se distinguent des définitions descriptives en ce qu'elles ne visent pas la pure et simple description d'une donnée. Le cas paradigmatique est celui de la création d'un mot, dont la définition révèle une intention de signification, ou celui d'une expression en usage pour laquelle est introduit un sens nouveau. Bien qu'il soit souvent difficile de retracer l'histoire de l'émergence des mots – notamment parce que leur appartenance au lexique se mesure à l'usage, qui admet des degrés – il ne semble pas déraisonnable de penser que l'introduction de concepts comme *surmoi*, *cyborg* ou *anthropocène* a pu résulter de l'énoncé de définitions stipulatives, de même que « groupe » en mathématiques, lorsque la communauté des algébristes opta pour cette dénomination de la théorie éponyme, après les travaux pionniers d'Évariste Galois. On distinguera ces exemples de l'acte par lequel Alexandre Flemming introduisit le mot « *penicillin* » pour désigner la substance qu'il avait isolée à partir du *Penicillium notatum* car il s'agissait alors de *désigner* une espèce naturelle, sans qu'il soit absolument évident qu'aucune définition satisfaisante ait pu d'emblée en être proposée.

À ces différents cas s'ajoute celui de la définition explicative, qui explicite, précise ou affine le sens que le dictionnaire ou l'usage confèrent à un mot, en ayant éventuellement recours à une distinction conceptuelle précédemment inaperçue ou non répertoriée, à une théorie nouvellement introduite, ou à quelque thèse philosophique. Sans être indifférente à l'usage, la définition explicative n'entend pas s'y conformer ou s'y limiter, et elle ne se réduit pas à la définition *analytique* dont il a été précédemment question (*supra*, § 3). Elle se présente en effet parfois comme le résultat d'un effort de clarification pour le sens d'un mot dont aucune définition connue n'est jugée satisfaisante, comme lorsqu'une philosophe explique ce qu'il convient d'entendre, selon elle, par « connaissance », « *Dasein* » ou « aliénation ». La définition qu'Einstein donne de la simultanéité, fondée sur une thèse relative à la vitesse de la lumière, offre un non moins bel exemple de définition explicative (Einstein 2012, ch. 9).

Ce terme désigne aussi l'explication conceptuelle, au sens que Carnap donne à l'anglais « *explication* », dont la traduction en français se révèle délicate dès lors qu'on souhaite éviter toute confusion avec « *explanation* ». Par une « explication », en ce sens, on entend la substitution d'un concept à un autre, jugé trop vague ou fluctuant pour l'usage théorique qu'on souhaite en faire, et auquel on préfère en conséquence un concept apparenté mais formulé en des termes plus précis, en ayant éventuellement recours à un lexique spécifique. En théorie des ensembles, la définition d'ω (oméga) comme le plus petit ensemble inductif offre en ce sens très particulier une définition explicative de l'ensemble des nombres entiers naturels. Dans ce cadre théorique, les nombres 0, 1, 2, 3… sont (ou du moins peuvent-être) définis par des ensembles : 0 par ∅ (l'ensemble vide), 1 par {∅}, 2 par {∅,{∅}}, 3 par {∅, {∅}, {∅,{∅}}}, etc. Les premiers, que Carnap nomme « *explicanda* » (ce qui

signifie, littéralement, qu'ils requièrent une explication) sont destinés, dans le cadre de certains usages théoriques où la substitution est utile, féconde ou éclairante, à être remplacés par les seconds, que Carnap nomme alors « *explicata* » (Carnap 2015, 47 ; 1997, § 2 ; 1962, 3). Une telle définition ne vise évidemment pas la synonymie de l'*explicandum* et de l'*explicatum*.

5. Définition et sens

Les philosophes et le sens commun s'accordent généralement à dire qu'une définition donne le sens du mot qu'elle définit. Ainsi Locke, pour qui « définir n'est autre chose que faire connaître le sens d'un mot (…) » (Locke 1983, livre III, ch. 4, § 6). Pourtant, s'il en était ainsi, il faudrait en conclure que l'intention de la célèbre définition de l'homme comme animal rationnel est d'identifier le sens des deux expressions, alors qu'à l'évidence elles n'ont pas le même *sens*. La difficulté, ici, ne réside pas uniquement dans l'absence de tout critère général de la synonymie ou le caractère éminemment énigmatique ou évanescent du sens de « sens ». Il semble assez clair au regard de l'usage que « homme » *ne signifie pas* « animal rationnel ». Tout au plus pourrait-on avoir la tentation de soutenir, en dépit des difficultés considérables que cette thèse soulèverait, qu'il se trouve que, de fait, tout homme est animal rationnel et tout animal rationnel est homme, c'est-à-dire qu'un individu quelconque est un homme si, et seulement si, il est un animal rationnel. Même en supposant que les deux concepts soient suffisamment bien circonscrits et que la précédente thèse soit vraie, ce qui est évidemment loin d'être acquis, il resterait *possible* que des animaux non humains soient rationnels. Or dès lors qu'on est prêt à admettre la *possibilité* d'animaux rationnels non humains, on voit mal comment « homme » pourrait *signifier* « animal rationnel ».

Ce qui est visé par l'examen de cet exemple est la distinction entre identité extensionnelle, identité nécessaire et identité de sens, ou synonymie. Affirmer qu'une définition donne le sens du mot qu'elle définit revient à poser que *definiens* et *definiendum* sont ou doivent être synonymes, ce qui n'est certainement pas vrai de tous les énoncés qu'on reconnaît communément comme définitions. Comment du reste formuler une telle exigence alors qu'on ne dispose d'aucun critère satisfaisant de la synonymie ? L'identité de 1 et du nombre de satellites de la Terre est extensionnelle (1 est, *de fait*, le nombre de satellites de la Terre) sans être nécessaire (le nombre de satellites de la Terre *pourrait* être différent de 1). On exige généralement davantage d'une définition que l'expression d'une identité extensionnelle (« le nombre du satellites de la Terre » pourrait difficilement passer pour une *définition* de « 1 »). Considérons maintenant deux énoncés : « 2 est le successeur du successeur de 0 » et « 2 est la somme de 1 et de 1 ». Ces deux identités peuvent être considérées comme nécessaires et l'une ou l'autre pourrait être posée comme définition de 2, mais comment affirmer que l'une ou l'autre exprime une relation de synonymie ? Le *sens* que l'usage confère à « 2 » est trop incertain pour en décider et l'on peut certainement définir 2 en suivant l'une ou l'autre voie sans pour autant avoir fait connaître par là le sens que l'usage donne à « 2 », sens qui ne saurait faire l'objet d'aucune décision.

Réussit-on à faire connaître le sens d'un mot en procédant par stipulation ? Dans le cas d'un mot nouvellement introduit, la définition est effectivement donatrice de sens mais dans ce cas, le rapport *definiens-definiendum* n'est pas synonymique puisque le *definiendum* n'a précisément, avant définition, aucun sens. La définition fait également connaître le sens d'un mot qui appartient déjà au

lexique si elle lui confère un nouveau sens au total mépris de l'usage – procédé incongru et peu recommandable qui consisterait par exemple à décréter que « 2 » signifiera « bouc-cerf ». Mais qu'en est-il si l'on procède par stipulation tout en cherchant, autant que faire se peut, à respecter l'usage, en définissant « 2 », par exemple, par « le successeur du successeur de 0 » ? Une telle définition, qui est l'effet d'une décision, est bien donatrice de sens, mais en donnant un sens stipulé, elle vise le sens que l'usage donne à « 2 » sans pouvoir prétendre le restituer en son intégrité ; elle fixe un sens par décret sans atteindre à la synonymie, car « la somme de 1 et de 1 », qui n'est certainement pas synonyme de « le successeur du successeur de 0 », aurait tout aussi bien pu convenir, compte non tenu de considérations pragmatiques ou méthodologiques qui pourraient faire la différence et nous inviter à opter pour l'une des deux définitions plutôt que l'autre. Une telle définition, qui a valeur explicative, n'en est pas moins stipulative, sans être arbitraire, et sans exprimer davantage aucune identité de sens.

Est-ce à dire qu'aucune définition n'exprime jamais aucune identité de sens ? Qu'en est-il de l'identité des synonymes ? Pourrait-on par exemple définir « madré » par « matois » ou « matois » par « madré » ? Abstraction faite d'éventuelles différences de connotation ou de niveau de langue (dont on pourra admettre, bien que cela soit évidemment discutable, qu'elles n'affectent pas l'identité de sens), une telle définition serait insatisfaisante aux yeux de quelqu'un qui ne maîtriserait le sens d'aucun des deux mots, mais également aux yeux de quiconque maîtriserait aussi bien l'usage de l'un que de l'autre. Dans ce cas, la reconnaissance d'une identité de sens n'aurait en tout cas pas valeur de définition *explicative*, circonstance qui pourrait exclure qu'elle soit reconnue comme une définition. Locke par exemple, dans sa définition de la définition, exclut l'identité synonymique : « définir n'est autre chose que faire connaître le sens d'un mot *par le moyen de plusieurs autres mots qui ne soient pas synonymes* ». D'où l'intérêt d'une distinction entre définitions abréviatives, stipulatives, descriptives ou explicatives, qui indique une variété de fonctions de l'acte définitoire.

On peut supposer que Locke admettrait en revanche que l'identité de sens puisse valoir comme *definiens* de la synonymie : « deux mots sont synonymes si, et seulement si, ils ont le même sens » ; ou, pour dépasser la simple équivalence extensionnelle : « que deux mots soient synonymes *signifie* qu'ils ont le même sens ». Cet énoncé pourrait-il valoir comme une définition de la synonymie aux yeux de quelqu'un qui reconnaîtrait, par sa maîtrise de l'usage, que « avoir le même sens » et « être synonymes » sont synonymes ? On pourrait soutenir que cet énoncé d'identité aurait alors pour lui valeur de *thèse*, ou de *vérité empirique*, non de *définition*, et qu'il n'aurait valeur définitoire qu'au regard de qui, ne maîtrisant que partiellement la langue, apprendrait par là ce que signifie « synonyme ». En un sens, on retrouve ici le problème des limites floues et de la variété des définitions, et admettre ou non qu'une telle thèse puisse avoir valeur de définition dépend d'une stipulation touchant le sens de l'acte de définir. Le principal intérêt de cet exemple est cependant ailleurs : il soulève la délicate et importante question de savoir si une définition peut exprimer une connaissance (ici, une vérité empirique) sans cesser du même coup d'être purement définitionnelle. Nous revenons sur cette question ci-dessous, au paragraphe 9.

6. Les indéfinissables

Il a souvent été noté qu'on ne pouvait tout définir et Locke, par exemple, justifie ainsi cette affirmation : « tous les mots ne peuvent point être définis, par la raison tirée du progrès à l'infini (…). Car où s'arrêter, s'il fallait définir les mots d'une définition par d'autres mots ? » (Locke 1689, livre III, ch. 4, § 5). L'argument se conçoit s'il est entendu qu'on s'interdira de procéder à la manière des dictionnaires, où rien n'exclut que la définition d'un mot ne fasse appel à des termes dont la définition a elle-même recours, directement ou par la voie d'une série de définitions intermédiaires, au mot en question ; si l'on s'interdit, en d'autres termes, toute définition circulaire ou toute série de définitions dont l'enchaînement finit par faire cercle. Mais comment justifier un interdit que le dictionnaire ignore ? La raison qu'invoque Locke est fondée sur une conception de la définition qui en fait le reflet de la composition des idées à partir d'idées simples, d'où l'affirmation selon laquelle « les idées simples ne peuvent point être définies et (…) ce sont les seules qui ne puissent l'être » (Locke 1689, livre III, ch. 4, § 6-7). Ainsi conçue, la définition ne se réduit pas au lien entre mot et sens ; elle procède également à l'analyse explicative du sens d'un mot.

La pensée de Pascal illustre une autre forme de fondationnalisme définitionnel lorsqu'il donne comme règles pour les définitions de « n'entreprendre de définir aucune des choses tellement connues d'elles-mêmes qu'on n'ait point de termes plus clairs pour les expliquer » et de « n'employer dans la définition des termes que des mots parfaitement connus ou déjà expliqués » (Pascal 1954, 596-597). La définition se doit d'éclairer la nature de la chose définie et l'on obscurcirait l'esprit autant que l'ordre des connaissances si l'on s'engageait dans une définition de l'absolument simple ou de ce qui se connaît de soi.

Le postulat philosophique d'idées simples, en soi indéfinissables ou « connues d'elles-mêmes », n'est cependant pas la seule raison d'exclure les définitions circulaires en reconnaissant que tout ne peut être défini. L'épistémologie des définitions soulève un problème distinct qui conduit à la même exclusion, celui d'un vocabulaire primitif restreint mais suffisant pour la langue tout entière. Plus précisément : on demande quel ensemble minimal de termes d'une certaine langue L peut servir de base à la définition des autres mots de L, sans supposer pour autant que cet ensemble doive être unique ni que ses éléments possèdent aucun caractère de simplicité ou d'indéfinissabilité absolu. Si l'on nomme « vocabulaire de base », ou « V_b », l'ensemble en question, ce qu'on demande est que tous les termes de L qui n'appartiennent pas à V_b puissent être définis à partir de ce vocabulaire. Si par principe d'économie une exigence de minimalité est adjointe, on demande en outre que les éléments de V_b soient indépendants, c'est-à-dire qu'aucun mot de V_b ne soit définissable à partir des autres mots de V_b. Dans la plupart des cas, on jugera préférable que les éléments de V_b soient, en un certain sens du mot, aussi « simples » que possible, mais comme une définition précise d'un tel caractère de simplicité est tout sauf évidente, cette contrainte additionnelle est, dans l'immédiat, laissée en suspens.

Cette formulation du problème demande à être précisée, car rien n'est encore dit des formes de définition admises ni des conditions qu'un énoncé doit satisfaire pour avoir valeur de définition. Avant d'en arriver à ces questions, considérons à titre d'exemple un domaine particulier de la connaissance, l'arithmétique, dans lequel la pratique des définitions est courante et relativement claire et où il n'est pas difficile d'isoler un vocabulaire de base V_b. Par exemple, si l'on dispose dans V_b de mots ou de signes pour 0, la propriété *être un nombre entier naturel*, l'addition (+), la multiplication ($\times$), la relation *inférieur ou égal* ($\leq$), et la fonction *successeur* (qui à tout nombre *n*

associe son successeur dans la suite des entiers naturels), on définit aisément, dans la langue usuelle, les autres mots communément usités en arithmétique, à commencer par les entiers positifs non nuls eux-mêmes (1, 2, 3, etc.), mais aussi les fonctions *carré*, *cube*, *exponentielle*, *factorielle*, *plus petit commun multiple*, *plus grand commun diviseur*, etc., les propriétés *être pair*, *être impair*, *être premier*, *être la somme de deux nombres premiers*, etc., les relations *être strictement inférieur à* (<), *être un multiple de*, *être congru à … modulo* …, etc. On pourra dire par exemple que par définition 1 est le successeur de 0, 2 est le successeur du successeur de 0, que le carré d'un nombre *n* est le produit de *n* par lui-même, qu'un nombre *n* est pair s'il existe un nombre *m* tels que *n* est égal au produit de *m* par 2, que deux nombres entiers *m* et *n* sont tels que $m<n$ si, et seulement si, $m\leq n$ et *m* n'est pas égal à *n*, etc. Le problème est de savoir si l'ensemble V_b proposé suffit à définir tous les autres termes en usage qui ressortissent spécifiquement de l'arithmétique, ce qui soulève une question préalable, celle de la langue dans laquelle ces définitions sont formulées.

7. Définition et logique d'arrière-plan

Dans l'exemple qui vient d'être donné, il est aisé de voir que les définitions font appel non seulement aux mots ou signes de V_b (à savoir 0, entier naturel, +, ×, ≤ et successeur) mais également à d'autres expressions qui ne sont pas propres au vocabulaire de l'arithmétique comme « de », « est », « il existe », « et » ou « ne pas ». Cela suffit à mettre en évidence l'équivocité du problème épistémologique posé, dont la solution pourrait en effet dépendre non seulement du choix de V_b mais aussi des ressources logiques et grammaticales du langage dans lequel les définitions sont formulées. À défaut de préciser ce point, le problème épistémologique de l'existence d'un V_b risque d'être vidé de son sens : s'interroger sur l'existence d'un V_b minimal et suffisant, qui concerne le lexique, suppose qu'une question corrélative soit soulevée au sujet de la grammaire.

Cette difficulté a pu longtemps passer inaperçue, tant qu'on n'imaginait pas d'autre *medium* possible pour la formulation d'une définition que les langues dites « naturelles », dont les ressources sont si étendues et les constructions syntaxiques si profuses et variées qu'il semble humainement impossible de formuler aucune règle qui, au-delà du pur et simple respect de la grammaire, permettrait de caractériser précisément les constructions linguistiques et logiques requises par telle ou telle définition et de préciser ainsi le sens du problème épistémologique de l'évaluation, de la mesure et de l'éventuelle limitation de ces constructions. *A contrario*, le problème se pose clairement dès lors que les ressources syntaxiques et logiques de la langue dans laquelle une définition est formulée sont précisément circonscrites, comme il arrive dans les langues formelles. Les restrictions possibles sont multiples. Pour n'indiquer qu'un exemple, les logiciens distinguent des langues du premier ordre, dans lesquelles les quantifications ne s'appliquent qu'aux *objets* du domaine considéré, et des langues du second ordre, dans laquelle elles s'appliquent également aux *ensembles* de tels objets. Si le domaine d'objets est un ensemble de nombres, la syntaxe d'une langue du second ordre autorise non seulement des constructions interprétables par « pour tout nombre *n* » mais aussi des formulations interprétées par « pour tout *ensemble* de nombres ». Or selon un théorème logique bien connu, si la structure des nombres entiers naturels n'est pas caractérisable au premier ordre, elle l'est au second ordre. Ce qui est définissable dans un langage est indéfinissable dans l'autre, dans lequel certaines constructions syntaxiques et logiques font

défaut, et cela pour le même vocabulaire spécifiquement arithmétique V_b. Le problème d'une détermination de ce qui est requis par une définition ne saurait donc être limité à la question du choix d'un lexique primitif et doit être étendu à la sélection du langage d'arrière-plan, et donc des ressources logiques auxquelles on s'autorise à faire appel. Exemple qui illustre les vertus clarificatrices des méthodes formelles, qui permettent de restreindre aisément les moyens logiques à l'œuvre dans une définition et de mesurer précisément les conséquences d'une telle restriction. Ici comme en bien d'autres cas, la logique nous enseigne que ce qui peut apparaître comme de petites différences à l'aune de la pure grammaticalité est susceptible d'entraîner, au point de vue épistémologique et logique, des conséquences radicales : tel concept, que telle construction logique rend accessible à la définition, est affecté sans elle d'un irréductible caractère d'indéfinissabilité.

Pour la définition des termes d'un langage L, on distingue généralement les signes ou expressions *logiques* (typiquement « et », « ne pas », « pour tout », « il existe », etc.) des signes ou expressions *non logiques* (par exemple, dans le cas d'un langage arithmétique, « 0 », « + », « × », « ≤ », « successeur », etc.). Deux problèmes radicalement différents doivent alors être distingués : d'un côté, comment définir les signes et expressions *logiques* de L, de l'autre comment définir les signes et expressions *non logiques* de L ? Le premier problème touche à la question d'une caractérisation des *constantes logiques* et à la possibilité même d'une distinction entre signes logiques et signes non logiques. La difficulté est d'autant plus redoutable que les notions à définir sont habituellement comptées au nombre des notions primitives par excellence, en sorte qu'on voit mal sur quelle base elles pourraient être elles-mêmes définies sans entrer dans un cercle que l'on cherche précisément à éviter. Cette aporie relève d'une littérature philosophique spécifique dont il ne sera pas davantage question ici mais sur laquelle on trouvera pléthore d'introductions, par exemple (Bonnay 2008, 177-203). Le second problème suppose en revanche comme donnée préalable un langage logique de base, auquel vient s'ajouter un lexique non logique posé comme primitif, notre ensemble V_b. Une nouvelle distinction s'impose alors entre, d'une part, le problème de la définition des éléments de V_b (comment donner un sens aux mots du vocabulaire considérés comme primitifs et, de ce fait, comme indéfinissables ?) et, d'autre part, le problème de la définition des mots de L qui ne sont pas dans V_b, à partir de V_b et d'un ensemble de constructions syntaxiques présupposées qui caractérisent la logique de base à laquelle on s'autorise à avoir recours.

8. L'éliminabilité du défini

Bien qu'on dise ou lise fréquemment que les définitions sont des abréviations, cela ne saurait valoir de toute définition, et en particulier de toute définition explicative. Il arrive néanmoins qu'une identité définitionnelle comme « 1 est le successeur de 0 » puisse être lue comme une définition abréviative, qui autorise la substitution de « 1 » à « successeur de 0 » et, à l'inverse, du second au premier. Une abréviation est introduite qui peut, en retour, être aisément éliminée.

La méthode d'abréviation par définition est néanmoins souvent plus complexe. Un exemple paradigmatique est celui d'un langage logique L dans lequel, pour des raisons de simplicité, les seuls connecteurs propositionnels retenus sont, par exemple, la négation ($\neg$) et le conditionnel ($\rightarrow$) alors même que, pour des raisons de commodité, l'on souhaiterait pouvoir faire également usage de la conjonction ($\wedge$) et de la disjonction ($\vee$) dans l'écriture des formules. La solution habituelle consiste à

dire dans un métalangage ML que pour toutes formules *A* et *B* de L, ($A \wedge B$) et ($A \vee B$) sont définies comme les abréviations respectives de $\neg$(A $\rightarrow$ $\neg$B) et de ($\neg$A $\rightarrow$ B) (le logicien aura noté ici plusieurs abus d'écriture : omission des guillemets et concaténation de signes de L et de ML, qui sont pourtant, en général, des langages différents ; de tels abus seront lus eux-mêmes comme des abréviations commodes). Les signes $\wedge$ et $\vee$ ne sont pas directement identifiés à des expressions auxquelles ils pourraient être substitués : au lieu de cela, on fait appel, dans le métalangage, à une circonlocution comme la suivante, dans le cas du signe de la conjonction :

pour toutes formules *A* et *B* de L, ($A \wedge B$) est une abréviation de $\neg$(A $\rightarrow$ $\neg$B).

Cette définition autorise la substitution de ($A \wedge B$) à $\neg$(A $\rightarrow$ $\neg$B), qui a vertu abréviative pour toutes formules *A* et *B*, mais n'est jugée recevable que parce qu'elle autorise la substitution inverse, de $\neg$(A $\rightarrow$ $\neg$B) à ($A \wedge B$), par laquelle le signe défini se trouve éliminé. Arnauld et Nicole, après Pascal, notaient aussi que le soin apporté aux définitions permettait d'« abréger le discours », et soulignaient l'exigence de pouvoir, à l'inverse, « substituer mentalement la définition [au sens du *definiens*] à la place du défini » (Arnauld et Nicole 1992, 82).

On se gardera d'en conclure que la condition d'éliminabilité soit limitée au cas des définitions abréviatives, car lorsque Quine affirme que « définir, c'est éliminer » ou que « définir une expression, c'est montrer comment s'en passer » (Quine 1992, 55), il ne soutient pas pour autant que la définition se réduise à une pure abréviation. À première vue, l'idée générale semble assez simple. Si « nombre pair » est défini par « nombre divisible par 2 », on doit pouvoir non seulement remplacer « 7 n'est pas divisible par 2 » par « 7 n'est pas pair », mais également procéder à la substitution inverse, par laquelle le mot défini se trouve éliminé.

Pourtant, l'exigence d'éliminabilité du terme défini est plus complexe qu'il n'y paraît. Tout d'abord, comme le montre l'exemple de la définition de la conjonction par la négation et le conditionnel, l'élimination du signe défini ne se réduit pas toujours à son pur et simple remplacement par le *definiens* : il arrive que la définition soit contextuelle. On ne saurait éliminer le signe $\wedge$ au profit d'une suite de signes posée comme identique à $\wedge$; l'opération d'élimination est différente et consiste à remplacer ($A \wedge B$) par $\neg$(A $\rightarrow$ $\neg$B), quelles que soient les formules *A* et *B*. Ce qui justifie la substitution n'est donc pas une identité entre $\wedge$ et une certaine suite de signes qui le définirait mais l'*équivalence* de ($A \wedge B$) et de $\neg$(A $\rightarrow$ $\neg$B), pour tous *A* et *B*. La condition d'éliminabilité du défini suppose donc, outre la définition elle-même, que l'on ait à disposition une notion d'équivalence entre énoncés. Si l'on convient de noter *s* le signe défini, L le langage sur la base duquel *s* est défini (langage auquel *s* n'appartient pas) et L' l'extension de L obtenue par adjonction de *s* à L et par définition de s, alors, que *s* soit éliminable signifie que pour tout énoncé *A* de L' dans lequel figure *s*, il existe un énoncé de L équivalent à *A*. Ou plus exactement : il existe un énoncé dont on puisse *prouver* qu'il est équivalent à *A*. La condition d'éliminabilité requiert non seulement une relation d'équivalence entre énoncés de L' mais également une méthode permettant de prouver que deux énoncés quelconques de L' sont équivalents.

Une complication supplémentaire surgit lorsque le principe d'éliminabilité est appliqué sans autre précaution, ce que l'exemple suivant laisse apparaître. Posons que la propriété *être premier*, pour un nombre entier naturel, se définisse par « être divisible par deux nombres exactement ». Ainsi, 2, 3, 5,

et 7 sont divisibles par 1 et par eux-mêmes mais par nul autre nombre et sont donc premiers. *A contrario*, 1 est divisible par un seul nombre, à savoir lui-même, et 4, 6 et 8 sont divisibles par plus de deux nombres, ils ne sont pas premiers. Cela étant, observons l'effet de l'élimination du défini sur la relation de conséquence logique, entre prémisses et conclusion.

« 13 est divisible par deux nombres exactement »

est assurément conséquence de

« 13 est premier »

sur la base de la définition précédente, par élimination du mot défini. En revanche, on ne saurait légitiment conclure

« Léa ignore que 13 est divisible par deux nombres exactement »

de

« Léa ignore que 13 est premier »

sur la même base, car rien de ne permet d'affirmer que Léa connaisse la définition de *premier* qu'on a donnée. De même, une application peu précautionneuse du principe d'éliminabilité du défini pourrait conduire à conclure faussement

« le successeur de 0 » est une expression formée d'un seul signe

du fait avéré que

« 1 » est une expression formée d'un seul signe.

Concluons de ces exemples que la condition d'éliminabilité du défini, qui est généralement requise des définitions, n'est en réalité nullement acquise si elle n'est assortie d'une précision relative aux *contextes* d'éliminabilité légitime. Le mot « premier » est éliminable au profit du *definiens* dans le contexte « 13 est… » sans l'être dans le contexte « Léa ignore que 13 est… », où le principe de substituabilité des identiques (en l'occurrence, il s'agit de l'identité définitionnelle de « premier » et de « divisible par deux nombres exactement ») n'est pas respecté. Cet exemple invite à s'interroger sur la nature de l'identité définitionnelle (entre *definiendum* et *definiens*) car la nature exacte de cette identité (identité extensionnelle, intensionnelle, ou identité de sens) conditionne également la condition d'éliminabilité.

9. Définition et conservativité

Une définition purement abréviative ne nous apprend rien, elle ne vise qu'à soulager la mémoire ou à réduire la longueur d'un discours qui, sans elle, pècherait par une insupportable prolixité. On a reconnu depuis longtemps que le remplacement effectif de « divisible », « premier », « pair », « carré », « cube », etc. par leur *definiens* en toutes leurs occurrences rendrait littéralement inintelligibles les énoncés arithmétiques. La concision du discours n'est cependant pas la seule vertu

attendue des définitions ; celles-ci ont également pour effet, et pour motivation, d'attirer l'attention sur une propriété, une relation, un nombre, une structure, qui présentent un intérêt suffisant pour mériter de passer sur les fonts baptismaux. L'efficacité abréviative ne suffit pas à expliquer l'invention de termes comme « sérendipité », « gyrovague » ou « anaclitique », dont l'introduction dans le lexique et la définition mettent en valeur des phénomènes dont l'importance – mathématique, médicale, sociologique, historique ou autre – est jugée telle par une certaine communauté de locuteurs qu'ils méritent d'être élevés au rang de concepts répertoriés dans la langue. Nul terme ne vient abréger de la sorte le concept *ardéchois célibataire né un 29 février* car nul ne voit l'intérêt d'en capter et retenir les contours par un nouveau morphème et sa définition. La saisie d'un concept par un mot enrichit les ressources lexicales en attirant l'attention sur un phénomène qui avait pu être précédemment remarqué ici ou là, sans que cet intérêt occasionnel fût allé jusqu'à cristalliser en un néologisme.

Pourtant, en tant que tel, ce genre de définition ne véhicule aucune connaissance, au sens *épistémique* du terme, qu'on pourra opposer à une connaissance *lexicale* ou *conceptuelle*. Je peux connaître le sens du mot sans savoir en aucune manière s'il existe une circonstance à laquelle il pourrait être appliqué ou s'il est applicable à tel ou tel objet. Or ce caractère purement conceptuel est souvent érigé en condition nécessaire de la définition : une pure définition n'a aucun contenu épistémique ; elle enrichit le registre de nos concepts sans nous rendre plus savants, en sorte que si un énoncé reconnu comme vrai augmente notre stock de connaissances, c'est que cet énoncé véhicule davantage qu'une pure définition. On parle, pour exprimer cette idée, de la *non-créativité* des définitions, ou d'un respect de la condition de *conservativité*.

Un énoncé peut enrichir mon vocabulaire en m'apprenant que par « tiercelet » on entend un oiseau de proie dont le mâle, comme chez l'épervier, est plus petit d'un tiers que la femelle. Mais un tel énoncé donne bien plus que la définition du mot « tiercelet » ; il m'apprend en outre, si j'ignorais ce fait, que l'épervier mâle est plus petit d'un tiers que la femelle, allégation que je dois tenir pour vraie dès lors que j'accepte la prétendue définition précédente. Et la connaissance épistémique dont l'énoncé en question m'instruit est exprimable dans la langue dont je disposais avant d'apprendre ce qu'est un tiercelet, comme un simple examen de la phrase « l'épervier mâle est plus petit d'un tiers que la femelle » suffit à le montrer. D'une telle « définition », on dira qu'elle ne respecte pas la condition de conservativité, ou de non-créativité, et qu'elle n'est donc pas *purement* définitionnelle.

Voici comment énoncer en termes plus précis cette condition : soit L un langage dans lequel est formulée une théorie T, c'est-à-dire un ensemble d'énoncés. Soit L' le langage obtenu à partir de L par adjonction d'un mot *m*, soit δ la définition de *m* formulée dans L' et T' la théorie obtenue en ajoutant δ à T. On dit que δ satisfait la condition de conservativité relativement à T si, et seulement si, tout énoncé de L démontrable à partir de T' est également démontrable à partir de T. Si T représente les connaissances dont on dispose et qui sont exprimables dans L, la condition de conservativité est satisfaite si, et seulement si, δ ne permet pas d'augmenter cet ensemble de connaissances. Que cette condition soit satisfaite est une condition nécessaire et suffisante pour que δ n'ait aucun contenu épistémique, pour que son apport soit purement conceptuel.

Le caractère créatif, ou au contraire non-créatif, d'un énoncé qui se présente comme une définition n'est cependant pas toujours aussi évident que dans l'exemple de « tiercelet » et la formulation des

définitions requiert quelques précautions. Imaginons une définition qui stipulerait que par « Alpha » on entend le plus grand nombre premier. De cette prétendue définition on déduit l'existence d'un plus grand nombre premier, ce qui contredit un célèbre théorème des *Éléments* d'Euclide. Cette pseudo-définition, dont le contenu devrait être purement conceptuel et n'impliquer aucune connaissance, contredit en fait des connaissances précédemment acquises ; elle ne satisfait pas la condition de conservativité. Plus généralement, introduire une notation pour désigner un objet défini par une propriété suppose que l'on s'assure au préalable qu'il existait bien un objet et un seul satisfaisant cette propriété, point sur lequel Frege a particulièrement insisté.

Il arrive que le non-respect de la condition de conservativité ait des conséquences dramatiques, dont le paradoxe de Russell offre un célèbre exemple. Une première version de ce paradoxe figure dans une lettre de Russell à Frege du 16 juin 1902 (Rivenc et de Rouilhan 1992, 240).

On se propose de définir un ensemble *E* par une propriété $\varphi(x)$ qui dépend de *x*. *E* est un ensemble d'ensembles. En symboles, la définition de *E* a la forme suivante :

$$\forall x\,(E{=}x \leftrightarrow \varphi(x))$$

ce qui signifie que *E* est l'ensemble qui satisfait la propriété φ (une difficulté logique est déjà présente dans cette prétendue définition : *E* n'est bien défini qu'à condition qu'il existe un tel ensemble). Cette propriété est elle-même définie ainsi : un ensemble *x* quelconque satisfait φ si, et seulement si, ses éléments sont les ensembles qui n'appartiennent pas à eux-mêmes. En symboles, $\varphi(x)$ est donc définie par :

$$\forall y\,(y \in x \leftrightarrow y \notin y).$$

La définition de *E* s'exprime donc par

$$\forall x\,(E{=}x \leftrightarrow \forall y\,(y \in x \leftrightarrow y \notin y))$$

d'où l'on déduit, par instanciation de *x* par *E* :

$$(E{=}E \leftrightarrow \forall y\,(y \in E \leftrightarrow y \notin y))$$

et donc, puisque *E*=*E* est une vérité logique :

$$\forall y\,(y \in E \leftrightarrow y \notin y)$$

Comme cela vaut de tout *y*, cela vaut de *E* en particulier, d'où à nouveau par instanciation :

$$E \in E \leftrightarrow E \notin E$$

En d'autres termes : *E* est élément de *E* si, et seulement si, *E* n'est pas élément de *E*, ce qui est contradictoire. La définition de *E* comme ensemble de tous les ensembles qui ne sont pas éléments d'eux-mêmes a pour conséquence une contradiction. Cette définition ne satisfait donc certainement pas la condition de conservativité et elle est même créative au plus mauvais sens du terme : elle engendre une contradiction. Ce genre de difficulté appelle l'énoncé de règles de la définition, assorties de conditions qui prémunissent contre de telles conséquences.

10. Règles classiques de la définition explicite

La conception classique des définitions, dont on trouve une illustration caractéristique chez Frege, requiert qu'elles satisfassent les conditions d'éliminabilité et de conservativité (Frege 1999, § 24). Dans un cadre formel, le respect de cette exigence est garanti par les règles de la définition, dont on trouve une exposition dans certains manuels de logique, par exemple l'*Introduction to logic* de Suppes (Suppes 1957, ch. 8). Celles que nous donnons ici s'appliquent aux trois catégories de signes non logiques des langues formelles pour la logique du premier ordre : symboles de relation, symboles de fonction et constantes d'individu.

Soient deux langues formelles L et L' pour la logique du premier ordre telles que L est incluse dans L' (en symboles : $L \subseteq L'$) ce qui signifie que tous les signes de L sont dans L' ou, de façon équivalente, que toutes les formules de L sont des formules de L'. On note L' – L (lire : L' moins L) l'ensemble des signes de L' qui ne sont pas dans L et on suppose que L et L' ont les mêmes signes logiques ; les éléments de L' – L sont donc des signes non logiques. Les règles de la définition donnent la forme d'une définition *explicite* des signes de L' – L sur la base de L. Chacune de ces définitions est un énoncé de L'.

Dans les exemples qui suivent l'énoncé de chacune des règles, L est le langage de la théorie des ensembles pure, dont le seul signe non logique est le symbole d'appartenance « $\in$ » : « $x \in y$ » se lit « x appartient à y », ou « x est élément de y ». Nous donnons successivement trois exemples de définition sur la base de ce langage :

Exemple 1 : définition du symbole « $\subseteq$ » pour la relation d'inclusion (« $x \subseteq y$ » se lit « l'ensemble x est inclus dans l'ensemble y ») ;

Exemple 2 : définition du symbole de fonction « P » pour les parties d'un ensemble (« $P(x)$ » se lit « l'ensemble des parties de l'ensemble x ») ;

Exemple 3 : définition du symbole « $\varnothing$ » pour l'ensemble vide (« $\varnothing$ » est une constante qui désigne un ensemble auquel rien n'appartient).

a. Règle de la définition explicite pour un symbole de relation

Soit R un symbole de relation à n places de L' – L. Ce qui est défini n'est pas le signe R lui-même mais l'expression $R(x_1,\ldots x_n)$, pour des $x_1, \ldots x_n$ quelconques. Cette formule atomique de L' (qui signifie que $x_1, \ldots x_n$ sont dans la relation R) est le *definiendum*. Le *definiens* est une formule φ de L dont les variables libres sont $x_1, \ldots x_n$.

La définition explicite de R par φ a la forme suivante :

$$\forall x_1 \ldots \forall x_n \ (R(x_1,\ldots x_n) \leftrightarrow \varphi)$$

où φ est une formule de L où ne figure aucune variable libre autre que $x_1,\ldots x_n$.

Exemple 1 : en théorie des ensembles, on peut définir le signe « $\subseteq$ » de l'inclusion par la formule

$$\forall x \forall y \ (x \subseteq y \leftrightarrow \forall z \ (z \in x \rightarrow z \in y \ )).$$

Cette formule exprime en symboles que pour tous *x* et *y*, *x* est inclus dans *y* si, et seulement si, tous les *z* qui appartiennent à *x* appartiennent aussi à *y*.

b. Règle de la définition explicite pour un symbole de fonction

Soit *f* un symbole de fonction à *n* places de L' – L. Dans la définition explicite de *f*, le *definiendum* n'est pas le symbole *f* lui-même mais la formule atomique $f(x_1,\ldots x_n)=y$ où *y* est une variable différente de $x_1,\ldots x_n$. La définition a la forme suivante :

$$\forall x_1 \ldots \forall x_n \forall y \ (f(x_1,\ldots x_n)=y \leftrightarrow \varphi \ )$$

où φ est une formule de L où ne figure aucune variable libre autre que $x_1,\ldots x_n$ et *y*.

Pour que la formule indiquée ait authentiquement la valeur d'une définition du symbole de fonction *f*, une condition supplémentaire doit néanmoins être satisfaite : il faut s'assurer que pour tous $x_1,\ldots x_n$ il existe un *y* et un seul tel que φ. Pour le dire en symboles, il faut pouvoir prouver

$$\forall x_1 \ldots \forall x_n \ \exists y \ \forall z \ (f(x_1 \ldots x_n)=z \ \leftrightarrow \ z=y).$$

Exemple 2 : on peut définir l'ensemble P(*x*) des parties de l'ensemble *x* par la formule (où il est fait usage de « $\subseteq$ », qui a été défini)

$$\forall x \forall y \ (P(x)=y \leftrightarrow \forall z \ (z \in y \leftrightarrow z \subseteq x)).$$

Cette formule est une manière de dire, en symboles, que les *z* qui sont éléments de P(*x*) sont exactement les *z* qui sont inclus dans *x*.

Pour que cette formule définisse effectivement le symbole de fonction P, il faut prouver que pour tout *x*, il existe un *y* et un seul tel que $\forall z \ (z \in y \leftrightarrow z \subseteq x)$ (preuve qu'il est possible de donner en théorie des ensembles).

c. Règle de la définition explicite pour une constante d'individu

Soit *c* une constante d'individu de L' – L. Dans la définition de *c*, le *definiendum* n'est pas le symbole *c* lui-même mais la formule atomique *c*=*y* où *y* est une variable. La définition explicite de *c* a la forme suivante :

$$\forall x \ (c=x \leftrightarrow \varphi \ )$$

où φ est une formule de L où ne figure aucune variable libre autre que *x*.

Pour que la formule indiquée ait authentiquement la valeur d'une définition de la constante d'individu *c*, il faut s'assurer qu'il existe un *x* et un seul tel que φ. Pour le dire à l'aide de symboles, il faut pouvoir prouver la formule

$$\exists x \forall y\ (c=y \leftrightarrow y=x).$$

Exemple 3 : on définit la constante ∅ pour l'ensemble vide par la formule

$$\forall x\ (\varnothing=x \leftrightarrow \forall y\ y \notin x\ ).$$

Pour que cette formule définisse effectivement la constante ∅, il faut pouvoir prouver qu'il existe un *x* et un seul tel que φ, c'est-à-dire ici tel que $\forall y\ y \notin x$ (ce qu'il est possible de faire en théorie des ensembles).

d. Définition explicite et signification

Dans chacune des règles précédentes, la définition explicite de *s* est une équivalence (précédée de quantificateurs universels) entre un *definiendum* (une formule atomique dans laquelle *s* est le seul signe non logique) et un *definiens* (la formule φ du langage L). Une telle définition ne donne pas à proprement parler la *signification* de *s*. Comment du reste le pourrait-elle si L est un langage non interprété ? En quel sens s'agit-il donc d'une *définition* de *s* ? La vertu d'une définition explicite de la forme indiquée est de déterminer l'interprétation de *s* en fonction de l'interprétation de L. Comme φ est une formule de L et que l'équivalence est précédée de quantificateurs universels, toute interprétation de L (au sens logique de l'interprétation d'un langage formel) détermine entièrement l'interprétation du *definiendum*, et donc de *s*. Encore faut-il comprendre que dans le cadre formel présupposé, l'équivalence du *definiendum* et du *definiens* est extensionnelle : il ne s'agit ni d'une équivalence nécessaire ni, *a fortiori*, d'une relation de synonymie.

11. Définissabilité implicite et explicite

a. Les définitions implicites au sens de Gergonne

Les règles énoncées au paragraphe 10 indiquent comment L peut être étendu en un langage L' par une définition explicite des signes de L' – L. Mais comment définir les signes de V_b, l'ensemble des signes non logiques de L, s'il s'agit justement des signes primitifs, considérés à ce titre comme indéfinissables ? La solution que proposa Joseph-Diez Gergonne au début du XIX^{e} siècle repose sur une distinction entre définitions implicites et explicites (Gergonne 1818-1819, 23). Elle trouve une illustration célèbre dans les *Fondements de la géométrie* de Hilbert (Hilbert 1971), où l'auteur commence par énoncer une liste d'axiomes formulés dans un vocabulaire de base très restreint (*point*, *droite*, *plan*, *passer par*, *entre* et quelques autres éléments de terminologie géométrique), avant d'enrichir par la suite le lexique du géomètre au moyen de définitions explicites. Aux yeux de l'auteur, ce sont les axiomes de la géométrie eux-mêmes qui valent comme *définitions* des éléments de V_b, à la différence de ce qu'on trouve dans les *Éléments* d'Euclide, où les énoncés premiers comprennent axiomes, postulats *et* définitions. Cette conception moderne de l'axiomatique

défendue par Hilbert fit l'objet des critiques de Frege, qui voyait dans le procédé hilbertien une regrettable confusion entre deux catégories d'énoncés qui, à ses yeux, devaient être soigneusement distinguées, les axiomes d'un côté, les définitions proprement dites de l'autre (Frege et Hilbert 1992).

Hilbert lui-même ne fait pas usage des termes « explicite » et « implicite » dans ce contexte mais sa méthode axiomatique fut ultérieurement intégrée à la théorie de la connaissance par des philosophes qui, tels Moritz Schlick, lui appliquèrent la terminologie introduite par Gergonne, popularisant ainsi dans la communauté philosophique une certaine conception des définitions implicites, également nommées « définitions par postulats » (Schlick 2009, § 7). Ces définitions prennent donc en fait la forme d'ensembles d'axiomes formulés dans un vocabulaire de base, qui contraignent les interprétations possibles du langage (car une interprétation du langage n'est alors admissible que si elle satisfait les axiomes) et contribuent ainsi à leur définition, non pas séparément et explicitement pour chaque signe de V_b, mais globalement et implicitement. D'un point de vue purement logique, rien n'exclut que ces axiomes soient des énoncés quelconques ; leur sélection est néanmoins habituellement censée refléter nos intuitions relatives au vocabulaire de base. Tel est effectivement le cas des axiomes que donne Hilbert dans les *Fondements de la géométrie* pour « point », « droite », « passer par », etc. et des axiomes de Peano pour les indéfinissables de l'arithmétique (« zéro », « nombre entier naturel », « successeur d'un nombre », etc.).

Notons T un ensemble d'axiomes formulés dans L qui peuvent être considérés comme une définition implicite du vocabulaire de base V_b. À moins que T soit contradictoire, plusieurs interprétations de L rendent vrais les énoncés de T. Idéalement, on pourrait souhaiter que ces axiomes soient si contraignants qu'une seule interprétation de L satisfasse T (c'est-à-dire soit modèle de T) ; l'interprétation du vocabulaire de base serait ainsi entièrement déterminée. On sait cependant que le mieux qu'on puisse espérer est que les modèles de T soient tous isomorphes et que ce n'est généralement pas même le cas. Selon un théorème de logique bien connu, le système d'axiomes de l'arithmétique connu sous le nom d'arithmétique de Peano (axiomes arithmétiques pour un langage du premier ordre) admet des modèles non isomorphes en sorte qu'il échoue à caractériser la structure des nombres entiers naturels. On voit que ce qu'on nomme généralement « définition implicite » du vocabulaire de base d'un langage par un système d'axiomes n'a force de définition qu'en un sens très relâché de ce terme. Encore faut-il préciser que ce qu'il est possible ou non de définir par postulats (c'est-à-dire, dans ce contexte, par des axiomes) dépend, ici aussi, de la logique d'arrière-plan dont on dispose. La structure des entiers naturels, définissable dans une logique du second ordre, ne l'est pas dans un langage pour la logique du premier ordre (cf. *supra*, § 6).

Exemple de l'arithmétique de Peano

Les axiomes de l'arithmétique de Peano (dans la logique du premier ordre) offrent un bel exemple de définition implicite. Pour un langage dont le vocabulaire de base est constitué de 0, s (la fonction successeur), + et ×, ces axiomes sont les énoncés suivants :

Axiome 1 : $\forall x \forall y$ (s(x)=s(y)→x=y) (si les successeurs de deux nombres sont égaux, ils sont eux-mêmes égaux).
Axiome 2 : $\forall x$ 0≠s(x) (0 n'est le successeur d'aucun nombre).
Axiome 3 : $\forall x$ (x+0=x) (définition de la somme d'un nombre et de 0).
Axiome 4 : $\forall x \forall y$ (x+s(y)=s(x+y)) (définition de la somme d'un nombre et du successeur d'un nombre).

Axiome 5 : $\forall x\ (x\times 0=0)$ (définition du produit d'un nombre par 0).
Axiome 6 : $\forall x \forall y\ (x\times s(y)= (x\times y) + x)$ (définition du produit d'un nombre par le successeur d'un nombre).
Axiomes 7 (schéma d'induction). Pour toute formule $\varphi(x)$ qui dépend d'une variable *x* (et qui définit ainsi un ensemble), on pose l'axiome suivant :

$$[\varphi(0) \wedge \forall x\ [\varphi(x) \rightarrow \varphi(s(x))]] \rightarrow \forall x\ \varphi(x)$$

qui peut se comprendre ainsi : pour tout ensemble (défini par la formule $\varphi(x)$), si d'une part 0 appartient à cet ensemble et si d'autre part, le successeur d'un nombre quelconque appartient à l'ensemble dès lors que ce nombre lui appartient, alors tout nombre appartient à l'ensemble.

Dans ce schéma d'induction, une quantification est énoncée dans le métalangage (« pour toute formule $\varphi(x)$ ») et ce schéma d'axiomes pose donc en fait une infinité d'axiomes. En logique du second ordre, à l'inverse, la quantification peut être formulée dans le langage lui-même et le nombre d'axiomes est alors fini. Cette différence est loin d'être purement stylistique. Dans le premier cas, en effet, chaque ensemble considéré doit pouvoir être défini par une formule du langage de la théorie, et il y a donc au plus un nombre *dénombrable* de tels ensembles ; dans le second cas, on quantifie en général sur un nombre *non-dénombrable* d'ensembles. Dans l'un et l'autre cas, la quantification n'a donc pas du tout la même portée.

b. La méthode de Padoa et le théorème de Beth

i. Définition implicite au sens de Padoa

Le cadre formel qu'on a décrit permet de construire un modèle – une représentation idéalisée – de ce qu'est une définition : étant donné un langage formel L dont les signes non logiques constituent le vocabulaire de base V_b, un ensemble T d'axiomes formulés dans L définit *implicitement* ces signes primitifs alors qu'un ensemble de définitions *explicites* enrichissent L en un langage L' plus étendu. Se pose alors la question du caractère minimal du vocabulaire de base, au sens de l'éventuelle définissabilité mutuelle de certains des signes de l'ensemble V_b qui a été choisi. Alessandro Padoa proposa la solution suivante de ce problème en introduisant une notion de définissabilité implicite, en un sens du mot « implicite » qui diffère de celui dont il a été question au paragraphe précédent. Soit *s* un signe de V_b et soit $V_b - \{s\}$ l'ensemble des signes de V_b différents de *s*. On dit que T définit implicitement *s* sur la base de $V_b - \{s\}$ (au sens de Padoa) si la condition suivante est satisfaite : chaque fois que deux modèles de T ont le même domaine d'individus et donnent la même interprétation des signes de $V_b - \{s\}$, ils donnent aussi la même interprétation de *s*. Autrement dit : la condition est satisfaite s'il n'existe pas deux modèles de T qui s'accordent sur $V_b - \{s\}$ sans s'accorder sur *s* ; ou encore : toute interprétation des signes de $V_b - \{s\}$ sur un domaine d'objets D détermine l'interprétation de *s* sur D.

ii. La méthode de Padoa

On distingue la définissabilité *explicite* de la définissabilité *implicite* (au sens de Padoa) de la manière suivante. Le signe *s* de V_b est *explicitement définissable* à partir de $V_b - \{s\}$ relativement à T s'il existe

une définition explicite (au sens des règles du paragraphe 10) de *s* à partir de $V_b - \{s\}$ qui est démontrable à partir de T. La définissabilité (explicite) d'un *definiendum* dépend de l'existence d'un *definiens* qui prend la forme précise d'une formule φ. On notera que cette conception resserrée de la définissabilité écarte la difficulté soulevée par le paradoxe de Berry (voir ci-dessus, § 2) en limitant le domaine du définissable à ce qui peut l'être par une formule φ d'un langage L bien déterminé.

Au vu des règles pour les définitions explicites, il n'est pas difficile de se convaincre que si *s* est *explicitement* définissable à partir de $V_b - \{s\}$ relativement à T, alors *s* est aussi *implicitement* définissable par T sur la base de $V_b - \{s\}$. Une application de ce fait est connue sous le nom de *méthode de Padoa* : pour prouver qu'un signe *s* n'est pas (explicitement) définissable à partir de $V_b - \{s\}$ relativement à T, il suffit de prouver qu'il n'est pas *implicitement définissable* par T sur la base de $V_b - \{s\}$ en exhibant deux modèles de T qui s'accordent sur $V_b - \{s\}$ et diffèrent en *s*.

iii. Le théorème de Beth

Il est plus difficile en revanche de prouver l'inverse, à savoir que si *s* est implicitement définissable par T sur la base de $V_b - \{s\}$, alors *s* est explicitement définissable à partir de $V_b - \{s\}$ relativement à T. Cette proposition, vraie pour les langages du premier ordre, est connue sous le nom de *théorème de définissabilité de Beth* (Boolos et Jeffrey 1989, ch. 24).

Il n'y a pas de théorème similaire pour les langages d'ordre supérieur, bien que les notions de définissabilité implicite (au sens de Padoa) et de définissabilité explicite puissent être aisément étendues à ces langages. Dans un langage du second ordre, rien ne garantit, en général, qu'un signe *s* implicitement définissable par une théorie T, sur la base de $V_b - \{s\}$ soit aussi explicitement définissable à partir de $V_b - \{s\}$ relativement à T.

c. Définissabilité d'un ensemble ou d'une relation

Les règles de la définition explicite valent pour la définition d'un signe dans le cadre formel précédemment décrit. Elles déterminent l'interprétation de ce signe pour toute interprétation du langage dans lequel est formulé le *definiens*. De même, une définition implicite au sens de Padoa vaut pour toute interprétation du langage dans lequel sont formulés les énoncés de T. On soulève un tout autre problème de définissabilité lorsque le langage L considéré est *interprété* dans un domaine d'objets D particulier, car ce qu'on définit alors est bien plus qu'un signe *s* pour une interprétation quelconque : le signe en question est alors interprété, et on caractérise donc un élément ou une partie de D, ou encore une relation (ou une fonction) sur D. De ce nouveau point de vue, on distingue deux notions de définition et deux sens de la définissabilité selon qu'elle se rapporte à une *structure* ou à une *théorie*.

i. Définissabilité dans une structure

L'un des problèmes de définissabilité les plus étudiés, en raison du caractère fondamental de la structure des nombres entiers naturels, est celui qui se rapporte à un langage L_A pour l'arithmétique, dont les signes non logiques sont, par exemple, « 0 », « s », « + » et « × », interprétés respectivement, dans l'ensemble N des nombres entiers naturels, par le nombre zéro, la fonction successeur, l'addition et la multiplication. Si l'on convient, par commodité, d'adopter la même notation pour les cinq signes indiqués et pour leurs interprétations respectives dans N, on pourra noter (N, 0, s, +, ×), ou $\mathcal{N}$ pour abréger, la *structure* définie par N, 0, s, + et ×. On demande alors si telle partie de N ou telle relation sur N est définissable dans la structure $\mathcal{N}$, ou, ce qui revient au même, dans le langage L_A, celui-ci étant considéré comme un langage *interprété*. Pour comprendre de quel sens de la définissabilité il est maintenant question, considérons quelques exemples.

- Le nombre 3 est définissable dans $\mathcal{N}$ par la formule « n= s(s(s(0))) ». Ce qu'on veut dire par là est que 3 est le seul nombre qui satisfait cette formule de L_A ; il est en effet le seul nombre qui soit identique au successeur du successeur du successeur de zéro.

- L'ensemble des nombres pairs est définissable dans $\mathcal{N}$ par la formule « $\exists m$ s(s(0))×m=n » de L_A, car un nombre n est pair si, et seulement si, il existe un nombre m dont le produit par le successeur du successeur de zéro est égal à n.

- La relation ≤ sur l'ensemble N est définissable dans L_A par la formule « $\exists k$ (k+m = n) », car deux nombres m et n sont tels que $m \leq n$ si, et seulement si, ils satisfont cette formule.

Étant donnés un langage L et une structure d'interprétation de L dont le domaine est D, une formule de L à une ou plusieurs variables libres définit un élément ou une partie de D, une relation ou une fonction sur D. Des questions de définissabilité similaires peuvent être posées pour des langages (interprétés) autres que L_A, et donc pour des structures autres que $\mathcal{N}$. La définissabilité d'une partie de D (ou d'un élément, d'une relation, etc.) dépend en général de la richesse du langage considéré.

Il est très aisé de voir que toutes les parties de N ne sont pas définissables (quelle que soit la richesse du langage pour l'arithmétique considéré) dès lors que l'on sait que l'ensemble des parties de N est indénombrable (d'après le théorème de Cantor, il n'existe pas de surjection de N dans l'ensemble des parties de N) alors que l'ensemble des formules d'un langage quelconque est infini mais dénombrable. Ce qui est plus difficile, et plus intéressant, est de chercher à caractériser précisément les parties de N et les relations sur N qui *sont* définissables par une formule d'un langage pour l'arithmétique, et quelle corrélation précise existe entre cette définissabilité et la richesse du langage considéré. La solution de cette question relève de la théorie de la calculabilité, où l'on montre par le détail que la définissabilité d'une partie de N dépend non seulement de la richesse en signes non logiques du langage L_A considéré, mais aussi et surtout des propriétés logiques de L_A (langage du premier ordre, du second ordre, etc.) et des formes logiques auxquelles on s'autorise à avoir recours (formules avec ou sans quantificateurs, avec ou sans alternance de quantificateurs, etc.).

ii. Définissabilité dans une théorie

Il existe un autre sens de la définissabilité, qui ne la rend plus, comme précédemment, relative à une *structure* mais, cette fois, à une *théorie*, par exemple à l'arithmétique de Peano (qu'on notera ici T_A) formulée dans un langage L_A pour l'arithmétique. Nous parlons ici encore d'un langage *interprété* dans un certain domaine d'individus D. Limitons-nous au cas de la définissabilité d'une partie de N (le domaine D est donc l'ensemble N des nombres entiers naturels) bien que la question se pose également pour celle d'un élément, d'une relation ou d'une fonction. Afin de prévenir toute confusion terminologique avec la définissabilité d'une relation *dans une structure*, on parle souvent de *représentabilité dans une théorie*, mais la question est bien celle de la possibilité de caractériser, et donc de définir, un ensemble par une formule de L_A. On dit qu'une partie *B* de N est définissable (ou représentable) dans T_A s'il existe une formule $\varphi(x)$ de L_A à une variable libre (à savoir *x*) telle que, pour tout entier nature *n*, les deux conditions suivantes sont réalisées :

$n \in B$ si et seulement si $\varphi(\dot{n})$ est démontrable dans T_A

$n \notin B$ si et seulement si $\neg\varphi(\dot{n})$ est démontrable dans T_A

où $\dot{n}$ est une représentation de l'entier naturel *n* dans L_A et $\varphi(\dot{n})$ est le résultat de la substitution de $\dot{n}$ à *x* dans φ. Par exemple, si *B* est l'ensemble des entiers naturels pairs, on peut montrer qu'il existe une formule $\varphi(x)$ (à savoir $\exists y\ s(s(0))\times y=x$) qui satisfait les deux conditions. L'ensemble des entiers pairs est définissable dans T_A. Pour les mêmes raisons que précédemment, il est aisé de voir que les parties de N ne sont pas toutes définissables dans T_A. La définissabilité d'un ensemble (ou d'une relation, d'une fonction, etc.) dans une théorie T dépend non seulement de la richesse du langage mais également de la force démonstrative de T : tel ensemble définissable dans telle théorie T n'est pas définissable dans une autre théorie T', pour un seul et même langage L.

d. Définissabilité d'une classe de structures

Si les *formules ouvertes* (avec variables libres) définissent des ensembles (ou des relations, etc.) dans une structure ou dans une théorie, les *énoncés* d'un langage formel (formules sans variable libres) ont également valeur de définition, mais en un sens différent. Considérons à titre d'exemple un langage formel L, non interprété, dont le seul signe non logique est un symbole de relation binaire R. Soit la théorie T définie par les trois formules suivantes de L :

$\forall x\ Rxx$

$\forall x \forall y\ (Rxy \rightarrow Ryx)$

$\forall x \forall y \forall z\ ((Rxy \wedge Ryz) \rightarrow Rxz)$

Ces formules expriment, respectivement, la réflexivité, la symétrie et la transitivité des relations binaires. Les relations qui satisfont ces propriétés sont des relations dites « d'équivalence ». Toutes les interprétations de L qui sont modèles de T (c'est-à-dire qui satisfont ces trois formules) interprètent le symbole R par une relation d'équivalence et, inversement, toute structure (D, R) où D est un domaine d'individus et R est une relation d'équivalence sur D est un modèle de T. La théorie T

définit la classe des structures d'équivalence et donc, en ce sens précis, la théorie T définit la notion de structure d'équivalence.

Afin d'éviter toute confusion terminologique avec un autre sens de la définissabilité (comme la définissabilité dans une structure ou dans une théorie) on parle habituellement plutôt de caractérisabilité ou d'axiomatisabilité mais il s'agit bien d'un nouveau sens de la définissabilité, ce qu'on voit mieux en posant le problème inverse : étant donné un certain langage formel L, dont le seul signe non logique est un symbole de relation binaire, la propriété, pour une relation binaire, d'être une relation d'équivalence est-elle *définissable* par un ensemble d'énoncés de L ? L'ensemble des trois formules précédentes donne une réponse positive : T *définit* le concept d'équivalence pour une relation binaire.

Une question similaire peut être posée pour d'autres propriétés, par exemple la propriété pour un ensemble d'être infini : la classe des ensembles infinis (considérée comme une classe de structures d'interprétations d'un langage L quelconque) est-elle définissable, ou caractérisable, par un ensemble d'énoncés de L ? Selon un théorème bien connu des logiciens, la réponse est que pour des langages du premier ordre, *être infini* est axiomatisable, mais uniquement par un ensemble T infini ; *être fini*, en revanche, n'est pas définissable, ou axiomatisable par un langage du premier ordre. L'étude systématique de la caractérisabilité des classes de structures (qui représentent des propriétés comme *être fini*, *être infini*, *être une structure d'équivalence*, *être une structure d'ordre*, etc.) relève d'une partie de la logique connue sous le nom de *théorie des modèles*.

12. Définition par abstraction

Dans *Les fondements de l'arithmétique* (1884), Frege discute d'un « mode de définition bien inhabituel » qui n'est cependant « pas totalement inconnu » (Frege 1969, § 63). Le procédé s'applique à des notions abstraites comme la *direction* (d'une droite) ou le *nombre* (d'objets qui tombent sous un concept) et consiste à choisir comme *definiendum* non la notion abstraite elle-même mais une relation d'identité. Par exemple, dans le premier des deux exemples indiqués, le *definiendum* est l'identité de deux directions de droites :

la direction de la droite *a* est identique à la direction de la droite *b* (pour deux droites *a* et *b*).

Le *definiens* est alors la relation de parallélisme entre *a* et *b*. En symboles, la définition de la direction (d'une droite) devient ce qu'on appelle un principe d'abstraction :

pour toutes droites *a* et *b* $\quad \mathrm{dir}(a) = \mathrm{dir}(b) \quad \leftrightarrow \quad a//\mathrm{b} \qquad (*)$

où « dir(*a*) » est une abréviation de « la direction de la droite *a* » et « *a*//b » une abréviation de « la droite *a* est parallèle à la droite *b* ». Ce qui justifie ce procédé de définition et le choix du *definiens* est que nous avons, selon Frege, une intuition géométrique du parallélisme de deux droites, mais nulle intuition claire du concept de la direction d'une droite, qui appelle donc une définition. À l'aide du principe d'abstraction (*), nous définissons ainsi une notion abstraite sur la base d'une relation connue par intuition. En définitive, Frege lui-même ne retient pas cette méthode pour sa définition

du nombre parce qu'il estime qu'elle ne nous donne aucune véritable lumière sur *ce qu'est* la notion abstraite définie, dans notre exemple, la direction de *a*.

Le principe d'abstraction s'applique plus généralement à des relations d'équivalence quelconques (sur la notion de relation d'équivalence, cf. *supra*, § 11 d). Soit R une telle relation sur un domaine d'objets D. R définit une *partition* de D, c'est-à-dire un ensemble de parties de D qui sont disjointes, non vides et dont la réunion est identique à D. Pour reprendre l'exemple précédent, le domaine d'objets pourrait être l'ensemble des droites de l'espace géométrique et R la relation de parallélisme entre ces droites. Les éléments de D se trouvent ainsi répartis et regroupés en sous-ensembles de D, dont les éléments sont tous similaires entre eux au regard de R. Dans notre exemple, chaque droite *a* de D définit une « classe d'équivalence » $[a]_{//}$ relativement à la relation de parallélisme entre droites. Pour chaque droite *a*, $[a]_{//}$ est l'ensemble des droites parallèles à *a*. Or si l'on fait abstraction des différences entre les droites d'une même classe d'équivalence, on ne retient que ce qu'elles ont en commun, à savoir une certaine propriété abstraite. Dans notre exemple, cette propriété est la direction de *a*, pour une droite *a* quelconque. Le principe d'abstraction (*) formulé ci-dessus pour la relation de parallélisme entre droites définit ainsi le concept de direction de droite. Dans les *Principles of Mathematics* (Russell 1903, § 109-110), Russell discute et critique une telle méthode de définition par abstraction. L'une des questions en jeu dans l'application de ce principe est de savoir si la définition permet d'affirmer que ce qui a été défini existe effectivement. Russell, quant à lui, adopte un principe d'abstraction qui évite précisément de s'engager sur l'existence des propriétés abstraites qu'il permet de définir. Plus d'un siècle plus tard, la valeur et les usages des différentes versions du principe d'abstraction font toujours l'objet de débats en philosophie des mathématiques car appliqué à certaines relations d'équivalence soigneusement choisies, ce principe offre une méthode de définition pour les structures de nombres, à commencer par celle des entiers naturels (Hale et Wright, 2005).

13. Vertus et vices des définitions circulaires

Une définition peut être qualifiée de *circulaire* si le terme à définir figure non seulement dans le *definiendum* mais également dans le *definiens*, ou si figurent dans le *definiens* des termes dont la définition fait appel, directement ou indirectement, au terme à définir. Ces définitions soulèvent un problème de légitimité, en particulier s'il n'existe aucune autre définition, non circulaire, du *definiendum*. Des raisons épistémologiques évidentes semblent justifier leur exclusion pure et simple : comment la définition pourrait-elle m'éclairer sur sens du *definiendum* si le mot à définir figure aussi dans le *definiens* ? Quelle valeur explicative lui accorder si l'explication fait appel à ce qu'elle est censée expliquer ? Et comment, par ailleurs, satisfaire la condition d'éliminabilité du défini si la définition d'une signe « *a* » repose sur un signe « *b* » dont la définition repose elle-même sur « *a* » ? Pourtant, le cercle d'une définition n'est pas nécessairement vicieux et il existe une variété de définitions circulaires parfaitement légitimes. Les définitions circulaires ne tournent pas nécessairement en rond et leurs vertus sont largement reconnues en mathématiques, en logique et en informatique, pourvu qu'elles satisfassent certaines conditions bien précises.

a. Définitions récursives

Dans le cas de la définition d'une fonction numérique *f*, la circularité peut constituer un obstacle au calcul des valeurs de *f*. Comment calculer *f*(*n*) et *g*(*n*) pour un entier naturel *n* quelconque si tout ce qu'on sait de *f* et de *g* est, par exemple, que pour tout entier *n*

$f(n) = g(n+1)$

$g(n) = f(n+1)$?

Si l'on tente de calculer *f*(0) en appliquant les clauses d'une telle définition, on entre dans un cercle sans fin qui ne conduit à aucun résultat :

$f(0) = g(1) = f(2) = g(3) = f(4)$ …

A contrario, la fonction *factorielle*, qui à 0 associe 1 et à tout entier positif *n* non nul associe le produit de *n* et des entiers non nuls inférieurs à *n*, est parfaitement bien définie par les clauses circulaires suivantes (où l'on note « fact(*n*) » la factorielle de *n*). Pour tout entier naturel *n*

$\text{fact}(0) = 1$

$\text{fact}(n+1) = (n+1)\times\text{fact}(n)$.

La circularité de cette définition est manifeste dans la seconde clause. Pourtant, une telle définition ne donne pas seulement une caractérisation générale de la fonction factorielle ; elle permet aussi le calcul effectif de sa valeur pour tout entier *n*, par application itérée de la seconde clause, *n* fois, suivie d'une application de la première. Ainsi, pour le calcul de fact(3) :

$\text{fact}(3) = 3\times\text{fact}(2) = 3\times2\times\text{fact}(1) = 3\times2\times1\times\text{fact}(0) = 3\times2\times1\times1 = 6$.

La circularité d'une définition n'est donc pas nécessairement vicieuse et les fonctions numériques dites « récursives » reposent même sur une version généralisée du procédé que l'on vient d'observer : le calcul d'une fonction *f* pour un certain argument *n* fait appel à la valeur de *f* pour un argument inférieur à *n*, jusqu'à atteindre un cas de base (0, dans l'exemple de fact), pour lequel l'évaluation n'est plus circulaire. Plusieurs questions se posent : à quelles conditions des clauses circulaires définissent-elles effectivement une fonction ? Comment identifier la fonction définie ? Comment savoir si elle est bien définie pour tout argument ? Comment mesurer sa complexité ? La *théorie de la calculabilité* donne des réponses à ce genre de problème de définition.

Une définition récursive de l'addition, qui permet de calculer *m*+*n* pour tous entiers naturels *m* et *n* est donnée par les deux clauses suivantes :

$m + 0 = m$

$m + s(n) = s(m+n)$

où s(*n*) est le successeur de *n*.

Cette définition, dont le caractère circulaire est également manifeste dans la seconde clause et qui permet le calcul effectif de la somme de deux nombres entiers naturels quelconques, est donnée

dans un langage *interprété*, où les signes « 0 », « s » et « + » ont leur signification usuelle. Elle se distingue, à cet égard, des axiomes d'une théorie pour l'arithmétique, non moins circulaires que les clauses précédentes, mais formulés dans un langage *non interprété* (dont l'interprétation peut donc être tout à fait différente de l'interprétation habituelle) et qui ont généralement, pour le symbole de fonction « + », la forme suivante :

$\forall x\ (x + 0 = x)$

$\forall x \forall y\ (x + s(y) = s(x + y))$.

En dépit de leur similitude avec les deux clauses précédentes, ces deux formules sont d'une tout autre nature ; elles ne constituent pas une définition récursive mais appartiennent au genre des définitions « implicites » d'un vocabulaire de base par une théorie, (cf. *supra*, § 11 a), genre dans lequel il n'y aurait aucun sens à vouloir bannir toute circularité puisque ce sont tous les éléments de V_b qui sont, globalement, implicitement définis par l'ensemble des axiomes.

b. Définitions inductives

Les définitions inductives sont une autre forme de définitions circulaires d'un usage très courant en logique et en mathématiques. La définition de l'ensemble des formules d'un langage L pour la logique propositionnelle est donnée par des clauses qui font elles-mêmes appel au mot « formule ». Le cercle apparaît clairement dans les clauses C_2 et C_3 :

C_1 : les lettres de proposition (p, q, r, etc.) sont des formules ;

C_2 : si φ est une formule de L, alors $\neg\varphi$ est une formule de L ;

C_3 : si φ et ψ sont des formules de L, alors $(\varphi \wedge \psi)$ est une formule de L.

Une telle définition est justifiée si l'on peut prouver que, sur un certain vocabulaire (formé de $\neg$, $\wedge$, et p, q, r, etc.) il existe un et un seul plus petit ensemble X tel que :

- p, q, r, etc. sont dans X
- si φ est dans X, alors $\neg\varphi$ est dans X
- si φ et ψ sont dans X, alors $(\varphi \wedge \psi)$ est dans X.

La théorie des ensembles nous apprend quelles conditions doivent être satisfaites pour que l'existence d'un tel plus petit ensemble X (l'intersection de tous les ensembles qui satisfont les trois conditions) soit garantie et permet de prouver, dans l'exemple de l'ensemble des formules de L, que cet exemple existe effectivement (Enderton 2001, 34-44). D'autres méthodes de définition, plus générales que celle qui vient d'être indiquée sur un exemple très simple, sont étudiées dans la théorie des définitions inductives, qui analyse les conditions auxquelles de telles définitions peuvent être justifiées (Aczel 1977).

c. Définitions prédicatives et imprédicatives

Une autre forme de circularité touche non l'usage des signes mais l'existence même des objets à définir. L'ensemble N des nombres entiers naturels est parfois défini comme le plus petit ensemble inductif, c'est-à-dire comme l'intersection de tous les ensembles X dont 0 est élément et qui sont clos par la fonction successeur (« clos » signifie ici que pour tout *x*, si *x* est élément de X, alors le successeur de *x* l'est aussi). Cette définition de N présuppose l'existence d'une collection, celle des ensembles inductifs, dont N est lui-même élément. L'objet défini appartient donc à une collection à laquelle il est fait référence dans le *definiens*. Ce procédé n'est guère problématique si l'on a, par ailleurs, l'assurance que la collection existe ; mais tel n'est pas toujours le cas, notamment s'il s'agit d'une collection infinie d'objets abstraits comme des propositions ou des ensembles et la définition peut alors s'apparenter à une forme de cercle vicieux. Pour qualifier une telle circularité, on parle de définition *imprédicative* ; à l'inverse, une définition est dite « prédicative » si elle ne comporte pas ce type de cercle.

Le paradoxe de Russell (cf. *supra*, § 9) repose sur une définition qui est clairement imprédicative. La recherche d'une solution conduisit Russell à formuler et invoquer le « principe du cercle vicieux », selon lequel la définition d'un certain *x* (objet ou propriété) n'est pas légitime si *x* figure dans le *definiens*, ou si le *definiens* comporte une quantification sur un domaine auquel *x* appartient (Russell 1906). Dans la version du paradoxe de Berry donnée ci-dessus (§ 2), par exemple, on prétend définir un nombre *b* qui ne peut être défini en moins de vingt-sept syllabes sur la base d'une totalité dont *b* fait partie. Poincaré rend compte du genre de paradoxe qui en résulte en rejetant lui aussi ce qu'il considère comme une définition vicieusement circulaire (Poincaré 1906, 307-310). En refusant toute définition imprédicative, on met cependant aussi à l'écart celles qui sont d'usage courant en mathématiques comme la définition des entiers naturels mentionnée ci-dessus, ce qui est loin d'être anodin.

Caractériser précisément les phénomènes d'imprédicativité et faire le départ entre les définitions imprédicatives qui engendrent des contradictions et celles qui sont non problématiques se révèle être extrêmement difficile, ce qui explique que les questions relatives aux définitions prédicatives et imprédicatives aient pu être source de réflexions profondes et de débats jusque dans la philosophie de la logique et des mathématiques contemporaines (Feferman 2005).

d. Définitions par révision

Est-il possible et utile d'assouplir les règles de la définition explicite (*supra*, § 10) en admettant que le signe à définir puisse figurer dans le *definiens* et de chercher ainsi à légitimer une nouvelle forme de définitions circulaires ? La théorie des définitions par révision apporte une réponse positive à cette question. Le moyen d'y parvenir est l'invention d'une nouvelle sémantique des définitions ; le prix à payer est l'abandon d'un des piliers de la théorie classique des définitions : la condition d'éliminabilité du signe défini.

Soit *D* un ensemble non vide considéré comme un domaine d'objets. Une formule φ à une variable libre d'un langage L interprété sur *D* définit un prédicat *F* dont l'extension est une partie *A* de *D*. La

définition explicite de *F* (et donc de l'ensemble *A*) est donnée par la formule $F(x) \leftrightarrow \varphi$ (précédée d'une quantification universelle : $\forall x$), où $F(x)$ est le *definiendum* et φ le *definiens*. Si l'on rompt la règle de non circularité des définitions explicites (cf. *supra*, § 10) selon laquelle *F* ne figure pas dans φ, on perd la garantie que *A*, qui est l'extension de *F*, soit bien défini. Par exemple, si φ est la formule $\neg F(x)$, on obtient une définition circulaire du *definiendum* $F(x)$, à savoir $F(x) \leftrightarrow \neg F(x)$, selon laquelle un élément de *D* satisfait $F(x)$ si, et seulement si, il satisfait $\neg F(x)$, ce qui, à première vue, semble absurde.

Les recherches de Belnap et Gupta sur la définition d'un prédicat de vérité (Gupta et Belnap, 1993) ont cependant mis en évidence l'intérêt de telles définitions circulaires, qui permettent de modéliser le comportement de certains énoncés paradoxaux bien connus comme l'énoncé dit « du menteur » (dont l'une des versions est « le présent énoncé n'est pas vrai »), qui doit être reconnu comme non vrai si l'on fait l'hypothèse qu'il est vrai, et comme vrai si l'on fait l'hypothèse qu'il ne l'est pas. Dans la théorie de la définition par révision, la sémantique des définitions circulaire est fondée sur l'idée d'une règle de révision, qui détermine l'extension du *definiendum* $F(x)$ non pas catégoriquement mais en fonction d'une valeur hypothétique qui lui est assignée dans le *definiens*. On étudie alors de façon systématique les phénomènes de variation ou de stabilité de l'extension de *F* dans le *definiendum* lorsque varie l'extension de *F* dans le *definiens* et lorsque l'application de la règle de révision est itérée un nombre fini ou infini de fois. L'hypothèse des auteurs de cette théorie est qu'elle ouvre la possibilité d'une définition de certains concepts, reconnus comme des concepts « circulaires ». Le concept de vérité ne serait pas le seul qui se révèlerait, à l'aune de cette méthode de définition, avoir un caractère circulaire (Gupta 2000).

14. Pourquoi définir ?

Les motivations de l'activité définitoire sont si multiples et si variées que les catégorisations classiques (définitions nominales ou réelles, implicites ou explicites, descriptives ou stipulatives, etc.) ne suffisent pas à couvrir ou rendre compte de tout le champ des pratiques de la définition. Celle-ci vise tantôt le recueil d'un usage au sein d'un groupe de locuteurs donné (« or », « argent »), tantôt la caractérisation d'une espèce naturelle sur la base de l'analyse d'un échantillon ou d'un exemplaire (le plomb, les tigres), tantôt celle d'une entité mathématique (le nombre π, l'ensemble des entiers naturels) ou encore la saisie de l'essence d'un universel (le Beau, le Juste), la clarification de notions générales par une doctrine philosophique (démocratie, connaissance), l'enrichissement du lexique par stipulation d'un sens attribué à des termes nouvellement introduits (« cyborg », « anthropocène »), la réforme d'un concept à la lumière d'une théorie scientifique nouvelle (simultanéité, travail), la compréhension précise de termes logiques fondamentaux (« et », « non », « pour tous »), ou encore la caractérisation d'une classe de structures mathématiques (les structures finies, les ordres totaux), pour ne rappeler que quelques exemples rencontrés au fil de l'analyse.

D'où la nécessité de distinguer, classer, clarifier les différentes formes, catégories et visées des définitions ; de formuler, aussi, les règles qui en régissent l'économie, afin de comprendre pourquoi certains énoncés donnent l'illusion d'authentiques définitions ou ne satisfont qu'en apparence les conditions et propriétés qui sont attendues d'elles (non circularité, éliminabilité du défini, non créativité, etc.). Les modèles offerts par les méthodes formelles mettent particulièrement bien en

lumière le sens de tels critères et les formes possibles de la définition, qu'elles rendent aussi accessibles à une précision toute mathématique. L'analyse logique de la définition renouvelle quant à elle les questions classiques soulevées par les philosophes de la tradition, ouvrant dans le même temps de nouveaux horizons tant pour notre compréhension des définitions des mathématiciens que pour la mise au point d'algorithmes capables de calculer les valeurs d'une fonction lorsque celle-ci est adéquatement définie.

Il serait cependant vain de chercher à fixer une fois pour toutes ce qu'est ou ce que doit être une définition, comme ce que sont ses formes et ses règles, car l'acte de définir est trop philosophiquement engagé pour pouvoir prétendre à la neutralité. Aussi la quête d'une essence conduit-elle à des définitions dont la nature et la fonction sont fort différentes si l'on compare par exemple les cas de Platon et d'Aristote. D'autres philosophes, qui rejettent l'idée même d'essence, préféreront quant à eux voir dans l'entreprise de la définition une analyse conceptuelle ou terminologique, fondée sur des méthodes empruntées à la logique ou sur l'examen du langage ordinaire. Sens et visée de la définition ne sont pas moins variables et sujets à interprétation dans le contexte des mathématiques, de la physique ou d'autres sciences. Dans tous les cas, la définition participe néanmoins d'un même idéal de clarification et donc, à ce titre, d'une activité qui mérite assurément toute l'attention du philosophe.

Bibliographie

Aczel, Peter, « Introduction to inductive definitions », in J. Barwise, éd., *Handbook of Mathematical Logic*, North-Holland, 1977, p. 739-782.

Aristote, *Seconds analytiques*, trad. fr. de P. Pellegrin, Paris, Flammarion, 2005.

Arnauld Antoine et Nicole Pierre, *La Logique ou l'art de penser* (1662), 5e éd. 1683, rééd. Paris, Gallimard, coll. Tel, 1992.

Auroux, Sylvain, « La définition et la théorie des idées », in J. Chaurand et M. Francière, *La définition*, Centre d'étude du lexique, Paris, Larousse, 1990, p. 30-40.

Bessé, Bruno de, « La définition terminologique », in J. Chaurand et M. Francière, *La définition*, Centre d'étude du lexique, Paris, Larousse, 1990, p. 252-261.

Belnap, Nuel, « On rigorous definitions », *Philosophical Studies*, 72, 1993, p. 115-146.

Bonnay, Denis, « Définir la logique, une question de simplicité », *Travaux de logique*, 19, p. 177-203, 2008.

Boolos, George S. et Jeffrey, Richard C., *Computability and Logic*, 3e éd., Cambridge University Press, 1989.

Buridan, Jean, *Summulae de dialectica*, livre 8, ch. 2 : « De definitionibus ». Trad. angl. de G. Klima, New Haven, Yale University Press, 2001.

Carnap Rudolf, « Les deux concepts de probabilité » (1945), trad. fr. de J. Boyer *in* Rudolf Carnap, *Logique inductive et probabilité, 1945-1970*, Paris, Vrin, 2015.

Carnap Rudolf, *Signification et nécessité* (1950) 2e éd. 1956, trad.fr. de F. Rivenc et Ph. de Rouilhan, Paris, Gallimard, 1997.

Carnap, Rudolf, *Logical Foundations of Probability*, Chicago, University of Chicago Press (1950), 2e éd. 1962.

Chapuis, André et Gupta, Anil, *Circularity, Definition and Truth*, Atascadero, CA, Ridgeview Publishing Company, 2000.

Chaurand, Jacques et Francière, Mazine, *La définition*, Centre d'étude du lexique, Paris, Larousse, 1990.

Czeżowski, Tadeusz, « On traditional distinctions between definitions », in T. Czeżowski, *Knowledge, Science, and Values*, Amsterdam, Rodopi, 2000.

Einstein Albert, *La théorie de la relativité restreinte et générale*, trad. fr. Paris, Dunod, 2012.

Enderton, Herbert B., *A Mathematical Introduction to Logic*, Academic Press, 2e éd. 2001.

Feferman, Solomon, « Predicativity », in S. Shapiro, éd., *The Oxford Handbook of Philosophy of Mathematics and Logic*, Oxford University Press, 2005, p. 590-624.

Frege, Gottlob, *Idéographie* (1879), trad. fr. C. Besson, Paris, Vrin, 1999.

Frege, Gottlob, *Les fondements de l'arithmétique* (1884), trad. fr. C. Imbert, Paris, Éd. du Seuil, 1969.

Frege, Gottlob et Hilbert, David, Correspondance, trad. fr. de J. Dubucs *in* F. Rivenc et Ph. de Rouilhan, éd., *Logique et fondements des mathématiques. Anthologie (1850-1914)*, Paris, Payot, 1992, p. 220-235.

Gergonne, Joseph-Diez, « Essai sur la théorie des définitions », *Annales de mathématiques pures et appliquées*, t. 9, 1818-1819, p. 1-35.

Gupta, Anil, « On circular concepts », in A. Chapuis et A. Gupta, *Circularity, Definition and Truth*, New Delhi, Indian council of philosophical research, 2000, p. 123-153.

Gupta, Anil, « Definitions », *The Stanford Encyclopedia of Philosophy*, édité par E. N. Zalta, été 2015, URL = <http://plato.stanford.edu>.

Gupta, Anil et Belnap, Nuel, *The Revision Theory of Truth*, A Bradford Book, 1993.

Hale, Bob et Wright, Crispin, « Logicism in the twenty-first century », in S. Shapiro, éd., *The Oxford Handbook of Philosophy of Mathematics and Logic*, Oxford University Press, 2005, p. 166-202.

Horty, John, *Frege on Definitions*, New York, Oxford University Press, 2007.

Joray, Pierre et Mieville, Denis, éd., *Définition. Rôle et fonctions en logique et en mathématiques*, *Travaux de logique*, 19, 2008.

Kant, Emmanuel, *Critique de la raison pure* (1781, 2e éd. 1787), trad. fr. d'A. Delamarre et F. Marty, in E. Kant, *Œuvres philosophiques*, t. 1, Paris, Gallimard, Bibliothèque de la Pléiade, 1980.

Kripke, Saul, *La logique des noms propres* (1972), trad. fr. de P. Jacob et F. Recanati, Paris, Éd. de Minuit, 1980.

Leibniz, « Méditations sur la connaissance, la vérité et les idées » (1684), trad. fr. de P. Schrecker in Leibniz, *Opuscules philosophiques choisis*, Paris, Vrin, 1959.

Leibniz, *Discours de métaphysique*, 1686.

Leibniz, *Nouveaux essais sur l'entendement humain*, 1765.

Locke, John, *Essai philosophique concernant l'entendement humain* (1689), trad. fr. de M. Coste, Paris, Vrin, 1983.

Neuwirth, Stefan, « Les définitions de nom et les autres », *Repères-IREM*, 100, 2015, p. 25-47.

Ockham, Guillaume d', *Somme de logique* (1323), livre I, trad. fr. de J. Biard, Mauvezin, T.E.R., 2e éd., 1993.

Pascal, Blaise, « De l'esprit géométrique et de l'art de persuader », in Pascal, *Œuvres complètes*, Paris, Gallimard, Bibliothèque de la Pléiade, 1954, p. 575-604.

Poincaré, Henri, « Les mathématiques et la logique », *Revue de métaphysique et de morale*, t. 14, n° 3, 1906, p. 294-317.

Quine, W. V., « Définition », in W. V. Quine, *Quiddités* (1987), trad. fr. de D. Goy-Blanquet et Th. Marchaisse, Paris, Éd. du Seuil, 1992, p. 55-57.

Robinson, Richard, *Definition*, Oxford, Clarendon Press, 1950.

Russell, Bertrand, « Lettre à Frege du 16 juin 1902 », trad. fr. in F. Rivenc et Ph. de Rouilhan, éd., *Logique et fondements des mathématiques. Anthologie (1850-1914)*, Paris, Payot, 1992, p. 240-241.

Russell, Bertrand, *Principles of Mathematics*, Cambridge, The University Press, 1903.

Russell, Bertrand, « Les paradoxes de la logique », *Revue de métaphysique et de morale*, t. XIV, n°5, 1906.

Schlick, Moritz, « La définition implicite », in M. Schlick, *Théorie générale de la connaissance*, 1918, 2e éd. 1925, trad. fr. de Ch. Bonnet, Paris, Gallimard, 2009, § 7, p. 77-86.

Suppes, Patrick, *Introduction to logic*, chap. 8: Theory of definition, New York, Van Nostrand Reinhold, 1957.

Pierre Wagner

pierre.wagner@univ-paris1.fr
Université Paris 1 Panthéon-Sorbonne, Institut d'histoire et de philosophie des sciences et des techniques