Publié dans Travaux de logique 16, 37-58, 2005,
source qui doit être utilisée pour toute référence à ce travail

Logicisme et définition explicite

Pierre Joray

Préambule

Présentant l'arithmétique comme une expansion définitoire d'un langage logique, le logicisme classique des *Principia Mathematica* fait curieusement peu de cas des définitions. Dès 1903, Russell écrivait:

> It is a curious paradox, puzzling to the symbolic mind, that definitions, theoretically, are nothing but statements of symbolic abbreviations, irrelevant to the reasoning and inserted only for practical convenience, while yet, in the development of a subject, they always require a very large amount of thought, and often embody some of the greatest achievements of analysis. (Russell 1903: 63).

Si elle ne reprend pas le mot de «paradoxe», l'introduction des *Principia* n'en avertit pas moins le lecteur quant au statut spécial des définitions:

> Theoretically, it is unnecessery ever to give a definition: we might always use the *definiens* instead, and thus wholly dispense with the *definiendum*. Thus although we employ definitions and do not define "definition", yet "definition" does not appear among our primitive ideas, because the definitions are no part of our subject, but are, stricly speaking, mere typographical conveniences. (Whitehead & Russell 1927: 11).

Plus sensibles au rôle et à la nature des définitions intervenant dans leurs constructions, les courants logicistes contemporains accordent généralement plus de place à une réflexion développée sur celles-ci. Pourtant si l'attrait pour des modes définitoires autres que la classique définition abréviative des *Principia* s'est trouvé au centre des réflexions – définitions implicites, inductives, par abstraction, par règles de Gentzen, ... – on ne trouve que peu de réflexions sur les possibilités effectivement offertes par les définitions explicites.

En dépit de l'existence d'une forte tradition issue de l'Ecole de Varsovie, consacrée aux définitions explicites, il est encore aujourd'hui très largement considéré que celles-ci se réduisent au mode abréviatif et conventionnel qui fut défendu par les auteurs des *Principia*. Souvent négligées, voire inconnues, d'autres formes explicites de la définition existent pourtant, dont la force d'accroissement expressif dépasse largement celle de la conception stérilisante de Whitehead et Russell.

Le but que nous poursuivons dans ces pages est de présenter l'une de ces formes renforcées de la définition explicite, celle qui fut développée dans les années mil neuf cent vingt par S. Leśniewski et A. Tarski. Et si, comme dans le cas des définitions implicites, nous verrons qu'elle permet un enrichissement effectif du formalisme où elle se trouve utilisée, nous justifierons son intérêt, en regard des définitions implicites, par l'avantage considérable qui est le sien: celui de garantir la consistance des expansions définitoires. Enfin, l'article se terminera par une réflexion sur le rôle d'un tel outil dans notre projet de logicisme catégoriel et sur la signification ainsi confèrée à la construction obtenue de l'arithmétique de Peano[1].

1. La définition dans les systèmes formels

Dans le cadre d'un système formel, on peut distinguer deux manières d'introduire un terme constant qui confèrent à celui-ci une position déterminée au sein de l'appareil déductif. La première est d'un usage incontournable puisque c'est par elle que sont inscrits les termes dits primitifs. Il s'agit de la caractérisation axiomatique. Le second mode d'introduction passe par l'ajout d'une définition. Cependant, une telle distinction est à bien des égards artificielle. D'un côté en effet il n'est

[1] Construction dont le lecteur trouvera le détail en fin du présent volume.

pas dénué de sens de considérer que la caractérisation axiomatique des termes primitifs constitue une *définition* de ceux-ci (qualifiée traditionnellement de *définition par postulats*). D'autre part, il est de nombreuses définitions dont on peut se demander si elles ne constituent pas, selon le mot de Łukasiewicz (1928b), des «axiomes masqués».

Dans ce domaine, comme dans bien d'autres, il convient de faire le départ entre les aspects objectifs du problème et ce qui, dans notre regard, revient à l'influence de la tradition et au poids de l'habitude. Il est ainsi usuel de ne parler de définition au sujet d'une expression que lorsque celle-ci introduit un terme unique et que ce terme apparaît comme nouveau ou supplémentaire par rapport à un langage donné ne contenant que des termes primitifs ou des termes préalablement définis. L'idée est alors que l'expression en question doit être à même de caractériser de façon univoque la signification du terme nouveau et ceci sur la base des possibilités expressives du langage où la définition se trouve inscrite.

Il apparaît alors que deux sortes de définitions sont susceptibles d'être acceptées comme telles. Il y a tout d'abord la définition *explicite*, celle que l'on qualifie volontiers de définition au sens propre du terme. C'est celle que l'on reconnaît comme posant qu'un certain *definiendum*, contenant le terme nouveau, signifie la même chose qu'une expression bien formée ne contenant que les anciens termes, le *definiens*. On dit aussi souvent qu'il s'agit des définitions par lesquelles le terme nouveau se trouve caractérisé sans qu'on ait besoin pour cela d'utiliser ce terme. Enfin, il y a aussi les définitions *implicites* et s'il n'est pas généralement possible d'y distinguer un *definiendum* et un *definiens*, cela est dû au fait qu'elles font usage du terme à définir pour le caractériser. Le qualificatif d'implicite revêt sans conteste une certaine connotation péjorative et il n'est pas inutile de rappeler qu'une immense majorité de logiciens adhèrent au principe méthodologique suivant:

> **Principe d'explicitation:** chaque fois que l'introduction d'un terme peut se faire par la voie d'une définition *explicite*, il convient de renoncer à la voie de la définition *implicite*.

Nous adoptons pleinement cette position, mais il nous faut aussi relever que, dans le cadre des langages formalisés, chacun des types de définition possède des avantages et des inconvénients propres. La définition implicite est réputée pouvoir donner lieu à des conséquences cachées, voire contradictoires, et sa véritable position au sein d'une construction formelle est considérée, à juste titre, comme celle d'un axiome additionnel. Ainsi l'ajout d'une définition implicite doit-il s'accompagner de toutes les précautions valant pour l'ajout d'axiomes. Malgré ces aspects «négatifs», la définition implicite est pourtant vue positivement comme un moyen puissant d'expansion des systèmes. A l'inverse, la définition explicite est jugée comme un moyen sûr et transparent d'introduction de termes nouveaux – on sait par exemple que, moyennant quelques précautions, elle ne peut avoir aucune conséquence contradictoire – néanmoins elle est souvent dépréciée comme un outil faible, purement conventionnel et dénué d'intérêt théorique propre.

Cette vue courante, comme nous allons le voir, est partiellement erronée et cela tient à la pauvreté de la conception standard de la définition explicite. Mais avant d'entrer dans la justification de cette affirmation, précisons ce que nous entendons par définition explicite.

> **Définition explicite.** Une définition est dite *explicite* losqu'elle pose dans une relation d'équivalence deux expressions distinctes (le *definiendum* et le *definiens*) et répond aux conditions suivantes:
>
> 1. Le *definiendum* contient une expression de la forme $\alpha(v_1, \ldots, v_n)$, où α est l'unique symbole à être défini et $v_1, \ldots, v_n$ sont des variables indiquant ses places d'arguments (s'il y en a).
>
> 2. Ni α, ni les variables $v_1, \ldots, v_n$ n'apparaissent dans le *definiendum* à plus d'une seule occurrence.
>
> 3. Le *definiens* est une expression bien formée du langage dans lequel la définition est posée; il contient uniquement des symboles préalablement introduits (en particulier, il ne contient pas α).
>
> 4. Toute variable (libre) apparaissant dans le *definiens* apparaît aussi dans le *definiendum*.

1.1 L'exemple de la négation

A titre d'exemple, considérons le système suivant, qui constitue une axiomatisation de la logique propositionnelle classique:

Système *L*

Alphabet:	$\{\supset, \sim, (,), p_1, p_2, \ldots, p_n, \ldots\}$	
Formules:	(a)	Tout p_i est une formule.
	(b)	Si α et β sont des formules, $\sim\alpha$ et $(\alpha \supset \beta)$ aussi.
	Rien n'est formule sinon par les clauses précédentes.	
Axiomes:	(A1)	$p_1 \supset (p_2 \supset p_1)$
	(A2)	$(p_1 \supset (p_2 \supset p_3)) \supset ((p_1 \supset p_2) \supset (p_1 \supset p_3))$
	(A3)	$(\sim p_1 \supset \sim p_2) \supset ((\sim p_1 \supset p_2) \supset p_1))$
Règles:	*Det* (détachement), *Sub* (substitution)	

Un tel système étant fondé et complet (en ce sens que l'ensemble des théorèmes coïncide avec celui des tautologies contenant $\supset$ et $\sim$), la théorie des ensembles adéquats de connecteurs, nous garantit que tout connecteur vérifonctionnel (unaire ou binaire) peut y être introduit par le biais d'une définition explicite abréviative. Par exemple, on pourra introduire conjonction et disjonction en posant:

$$P \wedge Q =_{\mathrm{df}} \sim (P \supset \sim Q)$$

$$P \vee Q =_{\mathrm{df}} \sim P \supset Q$$

Cela dit, relevons que la base axiomatique présente la particularité que la négation n'y apparaît que dans l'axiome A3. Celle-ci en effet est absente des deux premiers axiomes ainsi que des règles. Ce constat nous amène à remarquer qu'il est possible de considérer le système sous deux points de vue. D'un côté, on peut le considérer tel qu'il est présenté ci-dessus; on dira alors qu'il contient deux termes primitifs: la conditionnelle et la négation. D'un autre côté, cependant, il est aussi possible de considérer que l'on a affaire à une base plus faible qui ne contient que la conditionnelle comme connecteur primitif:

Système L^{abs}

Axiomes: (A1) $p_1 \supset (p_2 \supset p_1)$

(A2) $(p_1 \supset (p_2 \supset p_3)) \supset ((p_1 \supset p_2) \supset (p_1 \supset p_3))$

Règles: *Det* (détachement), *Sub* (substitution)

Ce qui était l'axiome A3 dans L devient alors une *définition implicite* de la négation dans L^{abs}:

(Déf$_{\sim}$) $(\sim p_1 \supset \sim p_2) \supset ((\sim p_1 \supset p_2) \supset p_1))$

La situation est cependant trompeuse. Affirmer que A3 peut jouer le rôle d'une définition implicite de la négation dans le système L^{abs}, c'est considérer que l'expression en question constitue une caractérisation de la négation basée sur la conditionnelle. Or tel n'est pas le cas, car sans la dite définition le système L^{abs} ne constitue pas une axiomatique complète de la logique classique de la conditionnelle. Il est connu que certaines tautologies conditionnelles ne sont pas théorèmes dans L^{abs}. Parmi celles-ci figurent en particulier les suivantes, la première étant la fameuse «loi de Peirce»:

(T1) $((p_1 \supset p_2) \supset p_1) \supset p_1$

(T2) $((p_1 \supset p_2) \supset p_2) \supset ((p_2 \supset p_1) \supset p_1)$

Ceci montre que, dans L, l'axiome A3 n'a pas pour *seul* rôle d'introduire la négation. Il constitue également une pièce maîtresse dans la caractérisation de la conditionnelle.

Le projet de faire de A3 une définition implicite de la négation n'est peut-être pas pour autant tout à fait inadéquat. Pour sauver l'idée, il convient seulement de remarquer que si l'on s'appuie non pas sur le système incomplet L^{abs}, mais sur un système complet de la conditionnelle, alors il devient possible de définir implicitement la négation. En ajoutant à L^{abs} la «loi de Peirce» à titre d'axiome, il est connu qu'on dispose d'un système conditionnel complet:

Système $L^{pos} = L^{abs}$ + T1

Axiomes:	(A1)	$p_1 \supset (p_2 \supset p_1)$
	(A2)	$(p_1 \supset (p_2 \supset p_3)) \supset ((p_1 \supset p_2) \supset (p_1 \supset p_3))$
	(T1)	$((p_1 \supset p_2) \supset p_1) \supset p_1$
Règles:	*Det* (détachement), *Sub* (substitution)	

Dans L^{pos} il est désormais possible de poser l'ancien axiome A3 comme une définition implicite de la négation, dont on peut affirmer qu'elle caractérise la négation sur la base de la conditionnelle[2].

1.2 L'explicitation de Lejewski

A ce stade de la réflexion, il convient de savoir qu'au lieu de s'appuyer sur A3 comme définition *implicite* de la négation, il est également possible, dans le système complet L^{pos} d'en donner une définition *explicite*, que le principe d'explicitation nous engage à préférer.

Le fait est peu connu et fut démontré par C. Lejewski (1958). La solution s'appuie sur une forme non élémentaire de définition explicite; une forme qui répond cependant point par point aux conditions spécifiques de la définition explicite. La définition de Lejewski peut être posée de différentes manières[3]. A fin de simplicité, celle que nous proposons ici s'inscrit par le biais d'un ajout au système des deux conditionnelles (tautologiques) suivantes:

(D1a) $(\sim p_1 \supset p_2) \supset ((p_1 \supset p_2) \supset p_2)$

(D1b) $((p_1 \supset p_2) \supset p_2)) \supset (\sim p_1 \supset p_2)$

Remarquons qu'ajouter D1a et D1b à titre de thèses dans un système complet de la conditionnelle, comme L^{pos}, revient à poser les expressions suivantes dans une relation d'équivalence déductive:

(*definiendum*) $\sim p_1 \supset p_2$

(*definiens*) $(p_1 \supset p_2) \supset p_2$

[2] Plutôt que A3, il est d'ailleurs possible de poser à titre de définition implicite de la négation l'expression plus faible: $(p_1 \supset \sim p_2) \supset ((p_1 \supset p_2) \supset \sim p_1))$.

[3] Pour les autres formes, cf. *infra*, section 2.

La déduction suivante montre qu'on dispose bien avec D1a et D1b d'une caractérisation adéquate de la négation classique:

1.	$\sim p_1 \supset \sim p_2$	*Prémisse*
2.	$\sim p_1 \supset p_2$	*Prémisse*
3.	$(\sim p_1 \supset p_2) \supset ((p_1 \supset p_2) \supset p_2)$	D1a
4.	$(p_1 \supset p_2) \supset p_2$	2, 3, *Det*
5.	$(p_2 \supset p_1) \supset p_1$	T2, 4, *Det*
6.	$(\sim p_1 \supset \sim p_2) \supset ((p_1 \supset \sim p_2) \supset \sim p_2)$	D1a, *Sub*
7.	$(p_1 \supset \sim p_2) \supset \sim p_2$	2, 6, *Det*
8.	$(\sim p_2 \supset p_1) \supset p_1$	T2 (*Sub*), 7, *Det*
9.	$((p_2 \supset p_1) \supset p_1) \supset (\sim p_2 \supset p_1)$	D1b, *Sub*
10.	$\sim p_2 \supset p_1$	5, 9, *Det*
11.	p_1	8, 10, *Det*

On obtient alors en effet l'ancien axiome A3, en appliquant deux fois le métathéorème de la déduction (disponible dans L^{pos}):

$$\sim p_1 \supset \sim p_2 , \sim p_1 \supset p_2 \vdash p_1$$

$$\sim p_1 \supset \sim p_2 \vdash (\sim p_1 \supset p_2) \supset p_1$$

$$\vdash (\sim p_1 \supset \sim p_2) \supset ((\sim p_1 \supset p_2) \supset p_1)$$

1.3 Un bilan

Méconnu, le résultat de Lejewski est pourtant instructif à plus d'un titre. Tout d'abord, et nous avons insisté sur ce point dans Joray (2005a), la notion standard d'*ensemble adéquat de connecteurs* s'avère largement dépendante de la conception usuelle de la définition. En s'autorisant à exprimer l'équivalence que doivent poser les définitions explicites par le biais des formules du système – ici en usant d'une double conditionnelle – Lejewski montre qu'il y a un sens en lequel l'ensemble $\{\supset\}$ est *adéquat*[4].

[4] Rappelons que dans la conception standard $\{|\}$ (barre de Sheffer) et $\{\downarrow\}$ (son dual) sont les seuls ensembles à connecteur unique qui soient adéquats.

Enfin, le résultat de Lejewski montre qu'en s'émancipant du cadre étroit fixé par Whitehead et Russell, il est possible de faire reculer significativement les limitations expressives usuelles de la définition explicite. L'exemple frappant de la négation, nous indique que, là où seule une définition implicite se trouvait envisagée pour l'introduction d'un terme, la voie d'une définition explicite reste parfois ouverte. L'étude de ces possibilités nécessite cependant de s'ouvrir à des modes non standard de définition explicite, tels celui de Lejewski et celui, plus élémentaire, que nous allons maintenant présenter.

2. La définition explicite comme thèse

Insatisfait de la faiblesse et du caractère non réglé des définitions qu'il rencontre dans les *Principia Mathematica*, Leśniewski envisage d'énoncer les définitions explicites en se passant du relateur non déclaré et métalinguistique «$=_{df}$». Avec l'aide de celui qui est alors son élève, A. Tarski, il développe une conception de la définition dans laquelle la relation d'équivalence posée entre le *definiendum* et le *definiens* se trouve exprimée à l'aide des seuls outils disponibles dans la base axiomatique utilisée: les termes primitifs. L'idée est simple: dès lors qu'une base axiomatique est suffisament forte pour exprimer le «si et seulement si» des définitions informelles, il devient possible de stipuler une équivalence entre un *definiendum* (D_{um}) et un *definiens* (D_{iens}) par le biais d'une (ou plusieurs) thèse(s) ajoutée(s).

En fonction du choix de tel ou tel ensemble de primitifs, diverses solutions peuvent être envisagées. L'idée d'utiliser à cette fin la biconditionnelle $\equiv$ est certainement la plus naturelle; les thèses définitoires à ajouter se coulent alors dans le shéma suivant:

$$\vdash\ D_{um} \equiv D_{iens}$$

Cependant, d'autres choix sont également possibles, s'appuyant par exemple sur la conditionnelle et la négation, la conditionnelle seule ou encore la barre de Sheffer. Les thèses définitoires prennent alors l'une des formes suivantes:

$\vdash \sim((D_{um} \supset D_{iens}) \supset \sim(D_{iens} \supset D_{um}))$ (avec ~ et ⊃ primitifs)

$\vdash D_{um} \supset D_{iens}$ et $\vdash D_{iens} \supset D_{um}$ (avec ⊃ primitif)

$\vdash ((D_{um} \supset D_{iens}) \supset ((D_{iens} \supset D_{um}) \supset p_i) \supset p_i$ (avec ⊃ primitif)[5]

$\vdash (D_{um} \mid D_{iens}) \mid ((D_{um} \mid D_{um}) \mid (D_{iens} \mid D_{iens}))$ (avec | primitif)

Notons que la troisième de ces formes est particulièrement intéressante puisqu'elle montre qu'il est possible de poser les définitions par le biais de thèses *uniques*, en s'appuyant exclusivement sur la conditionnelle. La seconde forme, qui procède par l'ajout de *couples* de thèses, est cependant plus aisée en pratique, même si elle enveloppe une conjonction implicite.

2.1 La créativité des définitions

L'une des particularités essentielles de la définition comme thèse est qu'elle se présente comme une procédure *interne* au système formel. Non seulement, il est désormais possible de poser les définitions en se passant du symbole métalinguistique *ad hoc* «$=_{df}$», mais, ce qui est plus remarquable, la façon d'inscrire les expressions à titre de thèses définitoires peut être entièrement réglée par une directive énoncée dans la base axiomatique. Ce faisant, les symboles définis ne constituent plus, comme dans les *Principia*, de simples convenances typographiques métalinguistiques, mais viennent enrichir officiellement le langage objet. Il y a dans l'usage des définitions comme thèses une dimension dynamique, puisqu'à l'inscription de toute nouvelle définition, l'ensemble des formules ainsi que celui des théorèmes se trouve élargi[6].

Cette caractéristique présente d'emblée un double avantage. Tout d'abord elle permet d'atteindre une formalisation plus complète de la démarche scientifique dans la mesure où la définition y joue souvent un rôle essentiel. D'autre part, elle répond à la conception intuitive que les définitions, dans le cadre scientifique, servent non seulement à

[5] Où p_i est une variable propositionnelle qui n'apparaît ni dans D_{um} ni dans D_{iens}.

[6] Cf. ci-après, section 2.4, notre exemple de système L^{def}.

abréger le discours, mais aussi à enrichir les possibilités expresssives du système linguistique utilisé.

Pourtant de nombreux logiciens ont préféré en rester, dans ce domaine, à la conception purement abréviative et externe de la définition. Outre la force de l'habitude, il y a plusieurs raisons qui permettent à notre sens d'expliquer un tel conservatisme. Tout d'abord, le fait que les définitions internes sont incompatibles avec la notion standard de système formel. Posé une fois pour toutes comme une entité close, entièrement déterminée par sa grammaire et sa base axiomatique, un système formel au sens usuel ne peut contenir de règle permettant un accroissement pas à pas des formules de son langage. L'adoption d'une conception interne de la définition nécessite en effet une refonte profonde – jugée parfois trop lourde – des notions de langage et de sytème formels[7]. Enfin, l'attachement à une conception statique des systèmes formels se double d'une méfiance traditionnelle vis-à-vis des possiblités déductives nouvelles que pourraient induire les définitions. La conception abréviative des définitions reste immanquablement liée à l'idée que les définitions ne doivent servir qu'à souligner ou rendre explicite ce que le système formel contient déjà et qu'elles ne doivent en aucune manière modifier sa force, c'est-à-dire ouvrir des voies déductives jusque là impraticables.

Une définition permettant des développements de ce genre dans le contexte formel de son inscription est dite *créative* et cette propriété peut être précisée de la manière suivante:

> Une définition D introduisant un symbole d dans un système S est dite **créative dans** S si et seulement s'il y a au moins une formule α de S ne contenant pas d, telle que α n'est prouvable dans S qu'avec l'aide de D (ou d'une autre définition).

L'introduction d'une définition dans un système est ainsi *créative* lorsqu'une formule indémontrable de l'«ancien» langage devient démontrable une fois la définition inscrite.

[7] On se référera sur ce point à la présentation très claire qu'en donne Miéville (2001-04), ainsi qu'à l'exemple de système avec règle de définition L^{def} que nous donnons ci-après, section 2.4.

2.2 Un exemple informel de créativité

Les auteurs des *Principia* l'affirment sans ambage, les définitions abréviatives externes ne modifient en rien le système où elles sont inscrites, elles n'ont donc, à strictement parler, aucune importance théorique. Sur la base d'une définition, il est en effet toujours possible de remplacer toute expression contenant le terme défini – expression forcément inofficielle – par une expression équivalente et officielle ne le contenant pas. On dit dans ce cas que les définitions garantissent l'*éliminabilité* des termes définis et l'idée est alors que s'appuyer dans la pratique sur un langage incluant ces termes à titre d'abréviations ne présente aucune espèce de problème: à celui qui contestera tel ou tel usage abréviatif, pense-t-on, on pourra toujours exhiber la formule officielle déployée ne contenant plus que des termes primitifs.

Ce dont Whitehead et Russell – et à leur suite nombre de logiciens – ne se sont pas aperçus, c'est que l'éliminabilité des termes définis n'est pas une condition suffisante pour garantir un usage «inoffensif» et sans conséquence des abréviations: même éliminable de toutes expressions un terme défini peut en effet avoir un effet *créatif* sur la conduite des preuves. Certes, comme nous l'avons montré dans Joray (2005c) aucun effet de ce genre ne peut avoir lieu dans les systèmes de logique propositionnelle usuels, mais le phénomène peut apparaître dès qu'un système formel possède une certaine complexité.

Pour le montrer, considérons le système propositionnel L^* ci-dessous. Il s'agit d'une expansion du système classique L (présenté plus haut dans notre section 1.1), obtenu, sans axiome supplémentaire, par l'ajout de variables de connecteurs binaires f_i ainsi que du quantificateur universel $\forall$ (qui sera utilisé uniquement pour lier les variables f_i). Plus précisément nous avons:

Système $L^* = L + \ldots$

Symboles ajoutés: $\{\forall, f_1, f_2, \ldots, f_n, \ldots\}$

Clauses additionnelles pour les formules:

(c) Si α et β sont des formules, $(\alpha f_i \beta)$ aussi.

(d) Si α est une formule, $(\forall f_i)\alpha$ aussi.

Règles additionnelles:

∀i Si une formule de la forme $\alpha \supset \beta$ est une thèse, alors $\alpha \supset (\forall f_i)\beta$ est également une thèse, pour autant qu'il n'y ait aucune occurrence libre de f_i dans α.

∀e Si une formule de la forme $\alpha \supset (\forall f_i)\beta$ est une thèse, alors $\alpha \supset \beta(f_i / \tau)$ est également une thèse, où τ est soit une constante de connecteur binaire soit une variable f_j libre pour f_i dans β et où $\beta(f_i / \tau)$ est le résultat de la substitution dans β de toutes les occurrences libres de f_i par τ.

Par construction, toute thèse de L demeure une thèse dans le système étendu L^*, par exemple les trois tautologies suivantes:

(T3) $p_1 \supset p_1$

(T4) $(p_1 \supset p_2) \supset ((p_1 \supset {\sim}p_2) \supset {\sim}p_1))$

(T5) ${\sim}{\sim}(p_1 \supset p_1)$

Les thèses spécifiques de L^* ne sont, par contre, pas même des formules de L. En voici un exemple, accompagné de sa preuve:

(T6) ${\sim}(\forall f_1){\sim}(p_1 f_1 p_1)$

1. $(\forall f_1){\sim}(p_1 f_1 p_1) \supset (\forall f_1){\sim}(p_1 f_1 p_1)$ [T3, *Sub*]
2. $(\forall f_1){\sim}(p_1 f_1 p_1) \supset {\sim}(p_1 \supset p_1)$ [1, ∀e, $f_1 / \supset$]
3. $(p_1 \supset p_1) \supset ((\forall f_1){\sim}(p_1 f_1 p_1) \supset (p_1 \supset p_1))$ [A1, *Sub*]
4. $(\forall f_1){\sim}(p_1 f_1 p_1) \supset (p_1 \supset p_1)$ [T3, 3, *Det*]
5. $((\forall f_1){\sim}(p_1 f_1 p_1) \supset (p_1 \supset p_1)) \supset (((\forall f_1){\sim}(p_1 f_1 p_1) \supset {\sim}(p_1 \supset p_1)) \supset {\sim}(\forall f_1){\sim}(p_1 f_1 p_1))$ [T4, *Sub*]
6. $((\forall f_1){\sim}(p_1 f_1 p_1) \supset {\sim}(p_1 \supset p_1)) \supset {\sim}(\forall f_1){\sim}(p_1 f_1 p_1)$ [4, 5, *Det*]
7. ${\sim}(\forall f_1){\sim}(p_1 f_1 p_1)$ [2, 6, *Det*]

Posons maintenant, à la manière des *Principia*, la définition abréviative suivante d'un nouveau connecteur binaire #:

$$(Def_{\#})\ \ P \# Q \ =_{df} \sim(P \supset Q)$$

En s'appuyant sur cette abréviation, on peut alors proposer la dérivation qui suit de la formule T7:

$$(\text{T7})\quad \sim(\forall f_l)\,(p_l\, f_l\, p_l)$$

1. $\sim(p_l \# p_l)$ [T5, *Def*$_\#$]
2. $(\forall f_l)\,(p_l\, f_l\, p_l) \supset (\forall f_l)\,(p_l\, f_l\, p_l)$ [T3, *Sub*]
3. $(\forall f_l)\,(p_l\, f_l\, p_l) \supset (p_l \# p_l)$ [2, ∀e, f_l / #]
4. $\sim(p_l \# p_l) \supset ((\forall f_l)\,(p_l\, f_l\, p_l) \supset \sim(p_l \# p_l))$ [A1, *Sub*]
5. $(\forall f_l)\,(p_l\, f_l\, p_l) \supset \sim(p_l \# p_l)$ [1, 4, *Det*]
6. $((\forall f_l)\,(p_l\, f_l\, p_l) \supset (p_l \# p_l)) \supset (((\forall f_l)\,(p_l\, f_l\, p_l) \supset \sim(p_l \# p_l)) \supset \sim(\forall f_l)\,(p_l\, f_l\, p_l))$ [T4, *Sub*]
7. $((\forall f_l)\,(p_l\, f_l\, p_l) \supset \sim(p_l \# p_l)) \supset \sim(\forall f_l)\,(p_l\, f_l\, p_l)$ [3, 6, *Det*]
8. $\sim(\forall f_l)\,(p_l\, f_l\, p_l)$ [5, 7, *Det*]

Cette dérivation semble parfaitement acceptable mais, contrairement à la dérivation précédente de T6, il ne s'agit pas d'une *preuve formelle* de T7, car les expressions contenant # ne sont pas des formules de L^*, mais seulement des abréviations de formules officielles plus longues. En de telles circonstances, on ajoute généralement qu'il suffit à chaque ligne concernée de remplacer l'expression incriminée par la formule officielle qu'elle abrège. Ceci se fait ici sans difficulté, pourtant la suite de formules ainsi obtenue ne constitue toujours pas une preuve. En effet, après l'élimination du terme défini #, la ligne 3 ne peut plus être justifiée comme le résultat de l'application de la règle ∀e à partir de la ligne 2:

2. $(\forall f_l)\,(p_l\, f_l\, p_l) \supset (\forall f_l)\,(p_l\, f_l\, p_l)$ [T3, *Sub*]
3. $(\forall f_l)\,(p_l\, f_l\, p_l) \supset \sim(p_l \supset p_l)$ [2, ∀e, f_l / ?] impossible!

Dans Joray (2005f), nous avons montré, par projection de L^* sur L, qu'il n'existe en fait aucune preuve de T7 dans L^*.

Cet exemple montre que, même dans un système relativement simple comme $L*$, un usage informel et non réglé de définitions abréviatives externes peut conduire à un renforcement des possibilités déductives du système – i.e. à un effet *créatif* de la définition puisque T7 ne contient pas # – et ceci bien que les termes définis soient parfaitement éliminables. On ne peut donc continuer à soutenir, avec les auteurs des *Principia*, que les définitions sont de pures commodités linguistiques.

De fait, deux réactions sont ici possibles. La première est une attitude de prudence: elle consiste à se restreindre systématiquement, dans la construction des preuves, à l'usage des termes primitifs. On peut cependant envisager la difficulté que représenterait la vérification que toutes les dérivations sont effectivement des preuves formelles dans une construction comme celle des *Principia Mathematica* ou par exemple dans une version formalisée de la théorie des ensembles de Zermelo-Fraenkel. Dans les deux cas (et dans bien d'autres que le lecteur n'aura pas de peine à trouver), le remplacement systématique des expressions abréviatives par des formules officielles ne contenant que des termes primitifs s'avèrerait très fastidieux puisqu'une majorité des preuves – et même certains axiomes – y sont exprimés à l'aide de termes définis.

Reste la seconde des réactions possibles. Plus audacieuse que la première elle consiste à «officialiser» le pouvoir créatif des définitions explicites en intégrant la procédure définitoire aux côtés des règles d'inférence usuelles de la base axiomatique. La chose ne va pas sans difficultés, mais elle fut examinée d'une manière méticuleuse par Leśniewski (1931). Il en résultat une conception non classique – dite *ouverte* ou *développementale* – des systèmes formels, dans laquelle un usage des définitions explicites, non contradictoire et entièrement réglé par la base axiomatique, permet un enrichissement pas à pas du langage formel et de certaines de ses possibilités déductives. C'est sur un système formel de ce genre que nous nous sommes appuyés pour notre construction logiciste et nous en donnons ci-après (section 2.4) un exemple propositionnel élémentaire. Mais avant d'y venir, il nous reste à examiner un fameux exemple de définition créative donné par

Łukasiewicz. Bien que fautif, comme nous allons le montrer, cet exemple reste important par l'influence qu'il exerça sur la méfiance qu'adoptèrent de nombreux logiciens à l'égard des définitions comme thèses.

2.3 Une erreur de Łukasiewicz

Comme on peut le constater dans (1928a) et (1928b), Łukasiewicz fut dès le début hostile à la conception des définitions que développèrent ses collègues Leśniewski et Tarski dans les années mil neuf cent vingt[8]. S'appuyant sur l'idée que tout mode définitoire donnant lieu à des possibilités créatives menait à l'adjonction d'«axiomes masqués», Łukasiewicz pensa être en mesure de donner un coup fatal à la conception leśniewskienne avec l'exemple de définition comme thèse qu'il publia dans son article sur la calcul équivalentiel (Łukasiewicz 1939). Dans cet article, une section se trouve en effet consacrée à la définition comme thèse, posée selon la méthode de Leśniewski, c'est-à-dire par l'ajout d'une thèse de la forme[9]:

$$\vdash \mathrm{E}D_{um}D_{iens}$$

Le contexte de l'exemple est le système biconditionnel assez étrange qui suit, basé sur une tautologie comme unique axiome et les trois règles: *détachement* (de la biconditionnelle E), *substitution* et une *procédure de définition*:

Système $\boldsymbol{Łuk^{Def}}$

Axiome:	$\mathrm{EE}s\mathrm{E}pp\mathrm{EE}s\mathrm{E}pp\mathrm{EE}pq\mathrm{EE}rq\mathrm{E}pr$
Règles:	*Det*, *Sub*, *Def*

Łukasiewicz montra alors que, si on s'interdit l'usage de la règle de définition *Def*, les seules thèses que l'on parvient à prouver sont des instances de substitution de l'axiome. En effet, la forme particulière

[8] Sur la polémique qui marqua les réflexions sur la définition dans l'Ecole de Varsovie, on se reportera à notre article Joray (2005d), qui contient une traduction française des comptes rendus de conférence de Łukasiewicz (1928a) et (1928b).

[9] Nous reprennons ici l'écriture préfixée de Łukasiewicz, où la lettre E désigne la biconditionnelle.

de celui-ci exclut de fait toute application de la règle *Det*. Puisque de nombreuses tautologies biconditionnelles ne sont pas des instances de substitution de l'axiome en question, il est clair que *Łuk*^{def} amputé de *Def* serait un système *incomplet* du calcul biconditionnel classique. En particulier, il est aisé de constater que les deux tautologies suivantes n'y seraient pas prouvables:

(*) E*pp*

(**) EE*pq*EE*rq*E*pr*

Notons que Łukasiewicz a montré au préalable que la tautologie (**) est suffisante, à titre d'axiome unique, et associée à *Det* et *Sub*, pour obtenir un calcul biconditionnel complet.

S'autorisant désormais à faire usage de *Def* dans *Łuk*^{def}, Łukasiewicz propose alors d'introduire la constante unaire V (le *verum*) par l'inscription de la thèse définitoire suivante:

(D) EV*p*E*pp*

Il montre alors, par la preuve qui suit, que (**) qui n'était pas prouvable sans *Def*, le devient à la suite de la définition D:

1.	EEV*p*E*pp*EEV*p*E*pp*EE*pq*EE*rq*E*pr*	[Ax. *Sub* (*s* / V*p*)]
2.	EEV*p*E*pp*EE*pq*EE*rq*E*pr*	[D, 1, *Det*]
3.	EE*pq*EE*rq*E*pr*	[D, 2, *Det*]

D'incomplet qu'il était sans usage de *Def*, le système *Łuk*^{def} devient donc complet par le pouvoir créatif de la définition D. Dans cette situation, soutient Łukasiewicz, la définition D constitue un axiome caché puisqu'elle vient modifier profondément les propriétés du connecteur primitif E. Clairement, D ne fait pas que caractériser le nouveau connecteur V, mais joue un rôle clef dans la caractérisation de E.

Łukasiewicz a parfaitement raison de soutenir qu'une telle force créative doit revenir à un axiome et ne peut pas rester le fait d'une simple définition. Cependant, il se trompe sur un point crucial qui aurait rendu D inacceptable aux yeux de Leśniewski: la définition D n'est pas une définition *explicite*, mais une définition *implicite*. La raison en est très simple: dans le contexte où D se trouve inscrite à titre de première définition, la tautologie (*) n'est pas dérivable. Il s'ensuit qu'en inscrivant D, la relation qui est posée entre le D_{um} et le D_{iens} n'est pas même *réflexive*, ce qui est une condition nécessaire à la relation d'équivalence exigée pour les définitions explicites.

D étant implicite, il n'y a donc rien de surprenant à devoir lui donner le statut d'axiome, mais rien ne dit avec cet exemple qu'il doit en aller de même des définitions posées comme thèses, lorsqu'elles sont véritablement explicites, comme le demandait Leśniewski.

2.4 Une logique avec directive de définition

Dans cette section, nous présentons un exemple simple de système développemental incluant une règle permettant de poser des définitions explicites par le biais de couples de thèses conditionnelles de la forme[10]:

$$\vdash \mathrm{C}D_{um}D_{iens} \quad \text{et} \quad \vdash \mathrm{C}D_{iens}D_{um}$$

Un tel système développemental a pour spécificité d'être construit pas à pas, par une inscription ordonnée de thèses. Les premières thèses de L^{def} sont constituées par ses trois axiomes[11]:

(α_1) $\mathrm{C}p\mathrm{C}qp$

(α_2) $\mathrm{CC}p\mathrm{C}qm\mathrm{CC}pq\mathrm{C}pm$

(α_3) $\mathrm{CCN}p\mathrm{N}q\mathrm{CCN}pqp$

Quant aux autres thèses, elles sont inscrites successivement par le moyen d'une des trois règles d'inférence de L^{def}: *substitution*, *détachement* et *définition*. Le système étant développé pas à pas, il convient que ses règles soient énoncées d'une manière qui prenne en compte l'*état* dans lequel celui-ci se trouve au moment de leur application. Un *état* du système se caractérise par l'ensemble ordonné fini des thèses déjà inscrites et il y sera fait référence par le biais de la

[10] Nous reprennons pour ce faire un mode de présentation utilisé dans Lejewski (1958).

[11] Ceux-ci sont similaires à ceux du système L présenté en section 1.1, mais exprimé ici dans l'écriture préfixée de Łukasiewicz, où C et N sont respectivement la conditionnelle et la négation.

dernière thèse inscrite. Par exemple l'état initial de L^{def} est la suite α_1, α_2, α_3, la dernière thèse inscrite étant alors α_3.

Afin de pouvoir énoncer les règles plus aisément, il est utile de se donner les conventions terminologiques suivantes:

- On nommera **constante** tout symbole consistant en une *lettre latine majuscule*, si besoin, indicée (dans l'interprétation visée, il s'agira d'une constante de connecteur).
- On nommera **variable** tout symbole consistant en une *lettre latine minuscule*, si besoin, indicée (dans l'interprétation visée, il s'agira d'une variable de proposition atomique).
- Les *lettres grecques* seront utilisées à titre de métavariables.
- Soit τ_i la dernière thèse inscrite d'un état de L^{def}. On qualifiera de **formule relative à τ_i** toute expression ε satisfaisant l'une des deux conditions suivantes: (1) ε est une variable; (2) ε est une expression de la forme $\Psi\omega_1\omega_2\ldots\omega_n$, où Ψ est une constante apparaissant dans au moins une des thèses de l'état concerné et $\omega_1\omega_2\ldots\omega_n$ sont *n* formules relatives à τ_i, *n* étant égal à 2 si Ψ est C, égal à 1 si Ψ est N et égal au nombre de variables différentes apparaissant dans la thèse où Ψ apparaît pour la première fois, si Ψ est une constante différente de C et N.

Sur la base de ces conventions métalinguistiques, on peut désormais expliciter les règles d'inférence comme suit:

Règle *Det*. Soit τ_i la dernière thèse inscrite d'un état de L^{def}. S'il y a parmi la liste des thèses inscrites dans cet état une thèse τ_j et une thèse de la forme $C\tau_j\varepsilon$, alors ε peut être inscrite comme nouvelle thèse à la suite immédiate de τ_i.

Règle *Sub*. Soit τ_i la dernière thèse inscrite d'un état de L^{def}. Soient enfin τ_j une des thèses inscrites de cet état et γ une formule relative à τ_i. Une expression ε peut être inscrite comme nouvelle thèse à la suite immédiate de τ_i si ε est obtenue en substituant γ dans τ_j à toutes les occurrences d'une certaine variable apparaissant dans τ_j.

Règle *Def*. Soit τ_i la dernière thèse inscrite d'un état de L^{def}. Un couple d'expressions ε_1, ε_2 respectivement des formes $C\Psi\,\omega_1\omega_2\ldots\omega_n\beta$ $C\beta\Psi\omega_1\omega_2\ldots\omega_n$ peut être inscrit à la suite immédiate de τ_i, pour autant que les conditions suivantes soient respectées:

(1) Ψ est constante différente de toutes celles qui apparaissent dans les thèses inscrites de l'état en question.

(2) $\omega_1\omega_2\ldots\omega_n$ ($n \geq 1$) sont n variables différentes les unes des autres.

(3) β est une formule relativement à τ_i.

(4) aucune autre variable que $\omega_1\omega_2\ldots\omega_n$ n'apparaît dans β.

A titre d'illustration, considérons maintenant un exemple d'état de L^{def} construit conformément aux axiomes et aux règles posées:

(τ_1)	C*Cp*CC*ppm*C*Cp*C*pp*C*pm*	[α_2, *Sub*]
(τ_2)	C*Cp*CC*ppp*C*Cp*C*pp*C*pp*	[τ_1, *Sub*]
(τ_3)	C*p*CC*ppp*	[α_1, *Sub*]
(τ_4)	CC*p*C*pp*C*pp*	[τ_2, τ_3, *Det*]
(τ_5)	C*p*C*pp*	[α_1, *Sub*]
(τ_6)	C*pp*	[τ_4, τ_5, *Det*]
(τ_7)	CN*p*N*p*	[τ_6, *Sub*]
(τ_8)	CA*pq*CN*pq*	[*Def* (1^e thèse)]
(τ_9)	CCN*pq*A*pq*	[*Def* (2^e thèse)]
(τ_{10})	CCN*p*N*p*A*p*N*p*	[τ_9, *Sub*]
(τ_{11})	A*p*N*p*	[τ_7, τ_{10}, *Det*]

Notons, au sujet de cette illustration, que la présence de τ_6 dans une situation qui n'est précédée d'aucune définition (et, en fait, la possibilité d'obtenir, dans une telle situation, toutes les tautologies contenant C et N) montre que l'erreur commise par Łukasiewicz est ici évitée: la règle *Def* conduit bien à poser des définitions *explicites*, exprimant une *équivalence* entre D_{um} et D_{iens}. On remarquera aussi que les thèses τ_8 et τ_9 définissent la constante A comme la disjonction. Enfin, le développement jusqu'à τ_{11} montre qu'il est possible de *construire* un état de L^{def} où l'expression usuelle du tiers exclu figure à titre de thèse.

3. Conclusion: une construction réglée de l'arithmétique

Plus simple à bien des égard que l'Ontologie – le système que nous avons utilisé pour notre logicisme catégoriel – le système L^{def} suffit pourtant à faire voir ce qui fait, dans notre perspective logiciste, la spécificité d'une logique développementale[12].

Tout d'abord, on notera que, contrairement aux systèmes formels standard, un système développemental ne s'appuie pas sur une caractérisation préalable de ses formules. L'ensemble des formules reste en effet constamment ouvert. La correction syntaxique des thèses inscrites au fur et à mesure du développement du système est à chaque étape assurée sur la base des thèses déjà inscrites – dans l'état initial, les trois axiomes – par les règles d'inférence qui jouent en fait également le rôle de règles de formation syntaxique. Le caractère contextuel qu'acquiert ainsi la syntaxe a pour conséquence qu'il est dénué de sens de demander, avant même avoir produit tel ou tel développement spécifique, si une certaine suite de caractères est ou n'est pas une formule ou une thèse de L^{def}. Une telle question ne peut avoir de sens que si elle se trouve relativisée à un état du système, état essentiellement caractérisé par les différentes définitions inscrites.

Ainsi, dans le cas du logicisme, il est dénué de sens de demander si telle ou telle représentation symbolique des propositions de Peano sont des thèses du système adopté comme point de départ. Le développement que nous appellons «logicisme catégoriel» est à comprendre comme une construction dont les modalités de progression sont entièrement gouvernées par les axiomes et les règles et qui montre qu'il existe un état effectif du système dans lequel des thèses correspondent aux propositions de Peano. La suite des définitions qui caractérise cet état et le travail du logicien qui consiste à les trouver

12 Sur les règles de définitions explicites de l'Ontologie, cf. ici même l'article de N. Gessler, section 1.3.

sont ici cruciaux, même au plan théorique. Le logicisme doit ainsi être compris non pas comme une *réduction* de l'arithmétique à la logique, mais comme une *construction* dont l'assise est logique et le développement entièrement réglé par la logique.

Enfin, il convient de relever la présence possible dans la construction de définitions créatives. Le lecteur qui se penchera sur la construction que nous présentons en fin d'ouvrage remarquera qu'à de nombreuses étapes, il est fait usage de définitions qualifiées d'«auxiliaires», en ce sens qu'elles ne sont introduites que dans le but de débloquer – ou à tout le moins de faciliter – la preuve de thèses qui ne contiennent pas les termes introduits par ces définitions. Sans doute, l'inscription de telles définitions s'apparente-t-elle à une sorte d'extension du système. Quoi qu'il en soit pourtant, l'inscription d'une définition explicite, même créative, se distingue de l'ajout d'un axiome ou d'une définition implicite sur deux points essentiels: étant explicite, la définition ne peut introduire de constante non logique; enfin, si elle est conforme aux règles de l'Ontologie, la définition explicite ne peut conduire à aucune inconsistance. Il a en effet été montré que l'Ontologie est un système consistant, en ce sens que ni sa base, ni aucun des développements qu'elle autorise, si riche soit-il en définitions, créatives ou non, ne peut envelopper d'inconsistance[13].

En regard des positions logicistes qui, comme par exemple le néo-frégéanisme, s'appuient sur une définition implicite comme le Principe de Hume, notre construction possède l'avantage que son existence à titre d'état de l'Ontologie constitue une preuve de consistance de l'arithmétique de Peano relative uniquement à celle de la logique utilisée[14]. Sans être réductionniste, notre construction donne bien ainsi lieu à un fondement de l'arithmétique qui peut être qualifié de logicisme.

[13] Sur la consistance des systèmes de Leśniewski, cf. Lejewski (1969).

[14] La construction néo-frégéenne en revanche n'assure pas *par elle-même* la consistance de l'arithmétique puisque, comme l'a montré Boolos (1987), la consistance de la logique du deuxième ordre augmentée du Principe de Hume n'est garantie que relativement à celle de l'arithmétique. Ceci montre que le logicisme néo-frégéen n'est pas indépendant d'autres approches fondationnelles et fait écrire à Hale et Wright que le Principe de Hume peut seulement «reasonably be regarded as consistent» (2001: 118).