Publié dans Travaux de logique 19, 135-157, 2008,
source qui doit être utilisée pour toute référence à ce travail

DÉFINITIONS EXPLICITES ET ABSTRACTION

Pierre JORAY

Préambule

En usant du *Principe de Hume* pour introduire par abstraction le nombre cardinal au sein du calcul des prédicats du deuxième ordre, les « néo-frégéens »[1] s'opposent curieusement à l'un des réquisits les plus centraux du projet logiciste de Frege : obtenir un édifice où chacun des termes arithmétiques se trouverait défini d'une manière explicite par les termes primitifs de la base logique. Sans doute convient-il, pour éviter les impasses du logicisme classique, d'affaiblir les conditions très exigeantes que Frege s'était imposées. Mais le risque encouru est que la démarche ne puisse plus être qualifiée de *logiciste* puisqu'elle s'appuie sur une définition non explicite du nombre, posée à titre d'axiome additionnel. En traitant manifestement le nombre comme un terme primitif, on compromet en effet le caractère éliminable de la définition, ainsi que la volonté chère aux logi-

---

1 Le projet néo-frégéen de C. Wright et B. Hale peut brièvement être décrit comme une réduction de l'arithmétique au calcul des prédicats du deuxième ordre s'appuyant sur une définition « implicite » du nombre cardinal. Cette définition, qui prend la position d'un axiome dans la construction, consiste en l'expression suivante, dite *Principe de Hume*, et dans laquelle « $\mathrm{N}x{:}Fx$ » doit être comprise comme signifiant « le nombre cardinal du concept F » et « $F \approx G$ » comme signifiant « il existe une relation biunivoque entre les objets tombant sous F et ceux tombant sous G » : $(\forall FG)(\mathrm{N}x{:}Fx = \mathrm{N}x{:}Gx \equiv F \approx G)$. Pour plus de détails, *cf.* Wright (1983), Hale & Wright (2001) et, dans la présente collection, Ebert & Rossberg (2007).

cistes de voir les termes arithmétiques déterminés d'une manière univoque par les constantes de pure logique[2].

Ainsi, nous apparaît-il préférable de considérer qu'une approche logiciste de l'arithmétique réclame l'usage exclusif de définitions explicites. Conserver une telle contrainte dans le projet logiciste que nous avons mené[3] n'est pas allé sans soulever certaines interrogations. En effet, si la définition explicite constitue une démarche sûre et pleinement transparente d'introduction des termes dérivés, elle est aussi souvent perçue comme un moyen trop faible, ne permettant aucune véritable expansion des théories. Depuis Whitehead et Russell, la définition explicite est conçue en logique comme un outil purement abréviatif et dénué d'importance proprement théorique.

Cette vue usuelle est pourtant erronée. Bien que largement ignoré des approches standard, cet état de fait était déjà connu dès les années vingt et nourrissait de riches discussions parmi les membres de l'Ecole de Varsovie. Les systèmes logiques de S. Leśniewski, faisant de la définition explicite un outil logique central et interne à l'axiomatique, en sont le meilleur témoin aujourd'hui. Ils constituent à nos yeux un cadre théorique adéquat pour fonder une entreprise logiciste.

Bien que se situant dans le cadre de notre travail relatif à l'arithmétique, cet article est spécifiquement consacré aux effets que l'ajout de définitions explicites peut avoir sur une théorie formalisée. Recherchés par certains, refusés par d'autres, les plus remarquables de ces effets sont ceux des définitions que Lukasiewicz qualifiait de « créatives », à savoir celles qui s'avèrent nécessaires à l'établissement de résultats ne mentionnant pas les termes définis. De Leśniewski, qui en fut le plus

2 Pour une discussion sur ce point, *cf.* Joray (2007).

3 Projet traité dans le cadre du Centre de Recherches Sémiologiques de l'Université de Neuchâtel, en collaboration avec N. Gessler et C. Degrange et avec le soutien du Fonds National Suisse de la Recherche Scientifique. *Cf.* dans la présente collection, Gessler, Joray, Degrange (2005) et Joray (2007).

ardent promoteur, il ne nous est guère parvenu de réflexions quant à l'enquête précise que nécessitait l'usage de telles définitions. Ses systèmes, ainsi que les directives définitoires méticuleusement formalisées qu'ils contiennent, nous sont en revanche connus[4] et ont permis de mettre en évidence qu'il est possible d'user des définitions créatives comme d'un outil remarquable d'expansion et de démonstration, tout en conservant les vertus attendues de la définition explicite (univocité, éliminabilité, non contradiction).

A notre connaissance, Leśniewski n'a jamais répondu à la critique de Lukasiewicz, qui considérait de telles définitions comme des axiomes cachés. Notre projet de construction logiciste de l'arithmétique ayant pour singularité de recourir à des définitions créatives, il nous fallait répondre à cette objection cruciale puisque la valeur de notre démarche en était pleinement dépendante. Cet article se propose d'enrichir les résultats que nous avons déjà publiés concernant cette problématique[5], en mettant en évidence que la créativité des définitions explicites ne repose pas sur l'introduction de significations nouvelles, mais sur leur dimension abstractive. Contrairement aux définitions « abréviatives », les définitions de Leśniewski ont en effet pour particularité remarquable de mettre à profit cette dimension. Il s'ensuit que les définitions autorisent le déclanchement de certaines possibilités déductives du langage formel concerné.

1. Les définitions explicites

Dans la mesure où les définitions explicites sont utilisées dans une grande diversité de contextes symboliques et avec la

4 Pour une présentation complète de ces systèmes nommés Protothétique, Ontologie et Méréologie, *cf.* Miéville (1984) et, dans la présente collection, Gessler, Miéville, Peeters (2001-07).

5 Joray (2005a), (2005b) et (2006).

visée d'introduire des termes de différentes catégories sémantiques – aussi bien des noms, des foncteurs que des connecteurs ou encore des opérateurs – il est assez délicat d'en donner une caractérisation générale qui ne dépende pas du cadre linguistique particulier où on entend y recourir. Par opposition à une définition « implicite » qui consiste à stipuler un énoncé ou schéma d'énoncé présentant un usage spécifiant d'un terme nouveau, on parle couramment de définition *explicite* lorsque le terme à introduire n'est pas lui-même utilisé en vue de déterminer la signification qui va lui être attachée. Plus précisément, on reconnaît une telle définition par le fait qu'elle pose comme équivalentes (ou comme ayant même signification) deux expressions propositionnelles distinctes dont la première – dite *definiendum* – introduit le terme à définir et la seconde – dite *definiens* – présente la signification qui va lui être associée, par le biais exclusif des «anciens» termes. Ainsi, dans les exemples suivants :

$$P \vee Q \quad =df \quad {\sim}P \supset Q$$

$$x \neq y \quad =df \quad {\sim}(x=y)$$

$$\mathrm{Inv}(R)(x,y) \quad =df \quad R(y,x)$$

$$x \in \cup(E,F) \quad =df \quad (x \in E) \vee (x \in F)$$

peut-on à chaque fois reconnaître des énoncés de la forme générale :

Definiendum $=df$ *Definiens*

où $=df$ exprime une relation d'équivalence entre un *definiendum* et un *definiens* qui devront répondre aux conditions suivantes :

1. Le *definiendum* contient une expression de la forme $\alpha(v_1,\ldots,v_n)\ldots(v_j,\ldots,v_k)$, où α (le symbole à définir) est

un symbole absent du langage dans lequel la définition est posée et $v_1,\ldots,v_k$ sont k ($k \geq 0$) variables indiquant les places des éventuels arguments de α.

On notera au sujet de cette première condition : a) que les arguments ne sont répartis en plusieurs parenthèmes que lorsqu'il s'agit de définir un foncteur formateur de foncteur (*cf.* notre 3[e] exemple ci-dessus) ; b) que la forme $\alpha(v_1,\ldots,v_n)\ldots(v_j,\ldots,v_k)$ constitue une partie seulement du *definiendum* lorsque la constante à introduire n'est pas d'une catégorie propositionnelle mais, par exemple, d'une catégorie nominale (*cf.* le 4[e] exemple ci-dessus).

2. α et les variables $v_1,\ldots,v_k$ apparaissent dans le *definiendum* à une et une seule occurrence.

3. Le *definiens* est une expression bien formée du langage où la définition est posée ; il ne contient que des symboles primitifs ou préalablement définis (en particulier, il ne contient pas α).

4. Le *definiens* ne contient comme variables (libres) que des variables qui apparaissent dans le *definiendum*.

Dans le cadre de cette caractérisation générale, il est possible de distinguer deux familles de définitions explicites. Eu égard à la manière dont se trouve exprimée la relation d'équivalence entre *definiendum* et *definiens*, nous parlerons respectivement de définitions *externes* ou *internes*.

1.1. Les définitions externes

Héritière des *Principia Mathematica*, la conception externe des définitions explicites est aujourd'hui encore la plus couramment pratiquée par les logiciens. En usant de métavariables telles que P, Q, … ainsi que d'un relateur métalinguistique usuellement noté *=df*, on peut par exemple définir le connecteur de disjonction $\vee$ sur la base du conditionnel $\supset$ et de la négation propositionnelle $\sim$. Pour ce faire, on posera par exemple l'expression suivante :

$$(\mathrm{Df}_{\vee}): \quad P \vee Q \quad =df \quad {\sim}P \supset Q$$

Il y a trois raisons pour lesquelles une définition de ce type mérite le qualificatif d'*externe*. La première est son caractère proprement métalinguistique. En effet, ni le relateur *=df*, ni les variables P et Q ne relèvent du langage objet, c'est-à-dire du langage que le symbole à définir doit venir enrichir. En outre, le fait que *=df* n'apparaisse pas parmi les primitifs du langage en question en font une constante qui n'est en aucune manière réglée par l'axiomatique fixant les significations logiques de base. Les auteurs des *Principia* s'expriment à ce sujet sans ambiguïté. Dans leur introduction, on lit en effet

> 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, strictly speaking, mere typographical conveniences. (Whitehead & Russell 1927, 11)

La seconde raison tient à ce que le terme défini ne se trouve aucunement intégré au langage pour lequel la définition est posée. De fait, le langage objet reste toujours le même ensemble de formules. Donné une fois pour toute par son vocabulaire et ses règles de formation, il n'est pas susceptible d'être élargi, sinon à être étendu à un langage distinct du premier. C'est ainsi que les termes introduits par définition ne sont disponibles que dans le métalangage et qu'ils ne font que donner la possibilité de désigner d'une manière abrégée les formules jugées parfois trop longues, et peut-être pas assez suggestives, du langage formel officiel.

Enfin, troisièmement, avec un tel outil, il demeure impossible de rendre compte dans le cadre strict d'une théorie formalisée de l'usage naïf qui est fait des définitions dans la version informelle de cette théorie.

Afin d'éclaircir ce dernier point, nous allons en donner une illustration dans le cadre simple du calcul propositionnel. Donnons-nous, pour commencer, une axiomatisation du calcul classique s'appuyant sur trois axiomes et deux règles d'inférence usuelles, le détachement du conditionnel (*modus ponens*) et la substitution :

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, alors $\sim\alpha$ et $(\alpha \supset \beta)$ en sont aussi.

Rien sinon n'est formule.

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* et *SUB*

Il est bien connu que le système L est *fondé* et *complet*, en ce sens que l'ensemble des thèses de L coïncide exactement avec l'ensemble des tautologies contenant le conditionnel et la négation. Les trois tautologies suivantes sont, par exemple, des thèses de L :

(T1) $p_1 \supset p_1$

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

(T3) $\sim\sim(p_1 \supset p_1)$

En outre, le système L est *fonctionnellement complet.* Toute fonction de vérité y est en effet exprimable et l'on peut choisir de mettre l'accent sur certaines d'entre elles en posant des définitions externes, par exemple, pour la conjonction, la disjonction ou encore pour une fonction moins usitée, à laquelle nous associons ici le symbole # :

(Df$\wedge$) $P \wedge Q =\text{df} \sim(P \supset \sim Q)$

(Df$\vee$) $P \vee Q =\text{df} \sim P \supset Q$

(Df#) $P \# Q =\text{df} \sim(P \supset Q)$

Considérons maintenant L^*, une expansion de L obtenue par l'introduction de nouvelles variables (les f_i, destinées à jouer le rôle de variables de connecteurs binaires), ainsi que du quantificateur universel $\forall$ (destiné à lier uniquement les variables f_i). Dans le détail, nous obtenons :

*Système L**

Symboles ajoutés :

$\{\forall, f_1, f_2, \ldots, f_n, \ldots\}$.

Clauses additionnelles pour les formules :

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

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

Règles additionnelles :

$\forall$i Si $\alpha \supset \beta$ est une thèse, alors $\alpha \supset (\forall f_i)\beta$ en est également une, pour autant que α ne contienne aucune occurrence libre de f_i.

$\forall$e Si $\alpha \supset (\forall f_i)\beta$ est une thèse, alors $\alpha \supset \beta\, [f_i / \tau]$ en est également une, où τ est une constante de connecteur binaire ou une variable f_j libre pour f_i dans β.

Clairement L constitue un sous-système de L^* en ce sens que toutes les formules et thèses de L restent respectivement des formules et des thèses de L^*. En dehors de ces formules communes aux deux langages, il y a évidement des thèses de L^* qui ne sont ni des thèses, ni même des formules de L. C'est le cas par exemple de la thèse T4 que nous donnons ci-dessous, accompagnée de sa preuve :

(T4) $\sim(\forall f_1)\sim(p_1 f_1 p_1)$

1.	$p_1 \supset p_1$	[T1]
2.	$(\forall f_1)\sim(p_1 f_1 p_1) \supset (\forall f_1)\sim(p_1 f_1 p_1)$	[1, *SUB*]
3.	$(\forall f_1)\sim(p_1 f_1 p_1) \supset \sim(p_1 \supset p_1)$	[2, $\forall$e, $f_1 / \supset$]
4.	$(p_1 \supset p_1) \supset ((\forall f_1)\sim(p_1 f_1 p_1) \supset (p_1 \supset p_1))$	[A1, *SUB*]
5.	$(\forall f_1)\sim(p_1 f_1 p_1) \supset (p_1 \supset p_1)$	[1, 4, *DET*]
6.	$((\forall f_1)\sim(p_1 f_1 p_1) \supset (p_1 \supset p_1)) \supset (((\forall f_1)\sim(p_1 f_1 p_1) \supset$ $\sim(p_1 \supset p_1)) \supset \sim(\forall f_1)\sim(p_1 f_1 p_1))$	[T2, *SUB*]
7.	$((\forall f_1)\sim(p_1 f_1 p_1) \supset \sim(p_1 \supset p_1)) \supset \sim(\forall f_1)\sim(p_1 f_1 p_1)$	[5, 6, *DET*]
8.	$\sim(\forall f_1)\sim(p_1 f_1 p_1)$	[3, 7, *DET*]

Si nous nous accordons à dire qu'une formule comme « $p_1 \supset p_1$ » exprime le caractère, disons, « répétitif » du conditionnel, alors la preuve que nous venons de voir est celle d'une formule qui peut être comprise comme niant qu'aucun connecteur binaire n'est répétitif (la formule exprime alors : *il existe au moins un connecteur répétitif*). On remarquera à la ligne 2 que la preuve s'appuie sur le remplacement de f_1 par le connecteur $\supset$ (connecteur qui est justement répétitif).

Mais revenons maintenant à la question des définitions explicites externes et considérons la définition déjà donnée plus haut du connecteur binaire # :

(Df#) $P \# Q =\text{df} \sim(P \supset Q)$

Sur la base de celle-ci, il nous est désormais possible de présenter la dérivation de la formule T5 suivante, dont nous pouvons remarquer qu'elle ne contient aucune occurrence de # et que nous pouvons comprendre comme niant que tous les connecteurs soient répétitifs (T5 exprimant alors : *il existe au moins un connecteur non répétitif*).

(T5) $\sim(\forall f_1)(p_1 f_1 p_1)$

1.	$\sim(p_1 \# p_1)$	[T3, Def#]
2.	$(\forall f_1)(p_1 f_1 p_1) \supset (\forall f_1)(p_1 f_1 p_1)$	[T1, *SUB*]
3.	$(\forall f_1)(p_1 f_1 p_1) \supset (p_1 \# p_1)$	[2, $\forall$e, f_1 / #]
4.	$\sim(p_1 \# p_1) \supset ((\forall f_1)(p_1 f_1 p_1) \supset \sim(p_1 \# p_1))$	[A1, *SUB*]
5.	$(\forall f_1)(p_1 f_1 p_1) \supset \sim(p_1 \# p_1)$	[1, 4, *DET*]
6.	$((\forall f_1)(p_1 f_1 p_1) \supset (p_1 \# p_1)) \supset (((\forall f_1)(p_1 f_1 p_1) \supset$ $\sim(p_1 \# p_1)) \supset \sim(\forall f_1)(p_1 f_1 p_1))$	[T2, *SUB*]
7.	$((\forall f_1)(p_1 f_1 p_1) \supset \sim(p_1 \# p_1)) \supset \sim(\forall f_1)(p_1 f_1 p_1)$	[3, 6, *DET*]
8.	$\sim(\forall f_1)(p_1 f_1 p_1)$	[5, 7, *DET*]

Si cette dérivation semble parfaitement acceptable, il convient pourtant de relever qu'au contraire de celle présentée plus haut pour T4, il ne s'agit pas d'une *preuve formelle* de T5. Dans un système comme L^*, une preuve formelle est nécessairement constituée d'une suite de formules du système en question. Or, puisque # n'est pas un élément du vocabulaire de L^*, les expressions des lignes 1, 3, 4, 6 et 7 de la dérivation ne sont pas des formules de L^*. Ces expressions, de fait, ne sont que des *abréviations* de formules officielles. Dans ce genre de situations, on rétorquera usuellement que la remarque est excessivement procédurière puisque, pour obtenir des preuves formelles officielles à partir de telles dérivations, il suffit d'éliminer les termes définis afin d'obtenir une suite de véritables formules de L^*. Les définitions explicites que nous avons examinées sont en effet éliminables dans le sens suivant :

> Une définition *D* d'une constante *d* est **éliminable** dans un langage *S* si et seulement si toute formule de *S* comprenant une ou des occurren-

> ces de d peut être montrée, sur la base de D, équivalente à une formule sans occurrence de d.

La définition de # est bien éliminable dans L^*. Pourtant, après élimination de toutes les occurrences du terme défini #, la suite obtenue ici ne constitue toujours pas une preuve formelle. On a bien alors à chacune des lignes une formule de L^*, mais c'est désormais l'application des règles d'inférence qui pose problème. Comme on le voit ci-dessous, la formule obtenue en ligne 3 ne peut plus être justifiée à partir de la ligne 2 par la règle $\forall$e, car aucun remplacement de f_1 dans « $p_1 f_1 p_1$ » ne peut conduire à « $\sim(p_1 \supset p_1)$ ».

2. $(\forall f_1)(p_1 f_1 p_1) \supset (\forall f_1)(p_1 f_1 p_1)$ [T1, *SUB*]

3. $(\forall f_1)(p_1 f_1 p_1) \supset \sim(p_1 \supset p_1)$ [2, $\forall$e, f_1 / ?] impossible !

De fait, on peut montrer qu'il n'existe aucune preuve de T5 dans L^* et donc que cette formule ne fait pas partie des thèses de ce système[6].

Au terme de ce détour, il convient de rappeler que nous entendions illustrer la troisième raison invoquée du caractère externe des définitions abréviatives standard, à savoir qu'elles n'offrent pas la possibilité de rendre compte formellement de l'usage naïf des définitions. En interprétant intuitivement les formules T4 et T5 comme exprimant respectivement « il y a au moins un connecteur binaire répétitif » et « il y a au moins un connecteur binaire non répétitif », nous avons implicitement considéré L^* comme formalisant une théorie naïve des connecteurs propositionnels, théorie dans laquelle, nous le savons, ces deux expressions sont vraies. N'ayant aucun statut au sein même du système formel L^*, la définition abréviative de # ne peut être

6 Pour la démonstration, *cf.* Joray (2005a, 193).

d'aucun secours dans la formalisation du raisonnement intuitif[7] qui mène à T5.

Plus grave peut-être est alors la confiance que nous plaçons dans les définitions abréviatives du fait de leur caractère éliminable. Notre exemple montre qu'une pratique peu attentive de ces définitions nous amène à considérer comme formellement démontrable une vérité qui en fait ne peut pas recevoir de preuve dans le système en question. Une fois cette illusion de complétude écartée par une plus grande attention, on pourra sans doute renforcer la base axiomatique de L^* de façon à obtenir un système où toute vérité de la théorie naïve soit démontrable.

Le choix méthodologique de Leśniewski fut différent. Il consista à estimer que la définition explicite devait être intégrée comme un moyen formel de démonstration des thèses. Autrement dit, il fit le choix d'une conception des définitions explicites comme *internes* aux systèmes formels.

1.2. Les définitions internes

Dans la conception interne des définitions telle qu'on la trouve chez Leśniewski, la relation d'équivalence entre le *definiendum* et le *definiens* doit être exprimée par des moyens propres au langage objet. L'usage d'un relateur métalinguistique comme *=df* se trouve alors proscrit au profit des seuls termes primitifs du langage adopté[8]. L'idée est simple. Dans un langage propositionnel comme *L*, dont les primitifs sont le conditionnel et la négation, on prouve que deux formules F_1 et F_2 sont *équivalentes* en montrant que les deux expressions suivantes sont des thèses :

7 Un tel raisonnement intuitif consisterait *grosso modo* en ceci : le conditionnel étant répétitif, le connecteur correspondant à la négation du conditionnel ne peut pas l'être et cela montre qu'il y a des connecteurs répétitifs et d'autres qui ne le sont pas.

8 Pour les questions liées à la définition dans les systèmes de Leśniewski, *cf.* les fascicules I et II de D. Miéville dans Gessler, Miéville, Peeters (2001-07).

$\vdash F_1 \supset F_2$

$\vdash F_2 \supset F_1$

L'idée de Leśniewski fut de munir la base axiomatique d'une règle permettant, sous certaines conditions bien précisées, d'inscrire à titre de thèses deux formules conditionnelles incluant un *definiendum* et un *definiens*, à savoir :

$\vdash$ *Definiendum* $\supset$ *Definiens*

$\vdash$ *Definiens* $\supset$ *Definiendum*

Dans le cas d'un système incluant le biconditionnel $\equiv$ à titre de primitif, la définition pourrait en outre être posée par l'inscription d'une unique thèse biconditionnelle[9] :

$\vdash$ *Definiendum* $\equiv$ *Definiens*

Bien que résultant d'un acte volontaire (choix d'un symbole nouveau et d'un *definiens* adéquat) les expressions définitoires sont néanmoins des thèses dans la mesure où elles sont obtenues par l'application d'une règle explicitée dans la base axiomatique[10].

Dans un système L^{*Def}, une version de L^* augmentée d'une règle définitoire à la Leśniewski, la dérivation de T5 commencerait par les lignes suivantes, dont les deux premières constituent la définition explicite de # et sont obtenues par application de la nouvelle règle :

9 C'est ainsi que la règle se présente dans les systèmes de Lesniewski et qu'elle se trouve aussi exploitée dans la thèse de doctorat de Tarski (1923). En outre, d'autres formes sont également possibles, *cf.* Lejewski (1958) et aussi Joray (2003).

10 On trouve une version méticuleusement formalisée de cette règle dans Leśniewski (1931). *Cf.* aussi la discussion et la présentation modernisée de cette règle dans Rickey (1975).

1. $(p_1 \# p_2) \supset \sim(p_1 \supset p_2)$ [*DEF* (a)]
2. $\sim(p_1 \supset p_2) \supset (p_1 \# p_2)$ [*DEF* (b)]
3. $(p_1 \# p_1) \supset \sim(p_1 \supset p_1)$ [1, *SUB*]
4. $(p_1 \supset p_1) \supset ((p_1 \# p_1) \supset (p_1 \supset p_1))$ [A1, *SUB*]
5. $(p_1 \# p_1) \supset (p_1 \supset p_1)$ [T1, 4, *DET*]
6. $((p_1 \# p_1) \supset (p_1 \supset p_1)) \supset (((p_1 \# p_1) \supset \sim(p_1 \supset p_1)) \supset \sim(p_1 \# p_1))$ [T2, *SUB*]
7. $((p_1 \# p_1) \supset \sim(p_1 \supset p_1)) \supset \sim(p_1 \# p_1)$ [5, 6, *DET*]
8. $\sim(p_1 \# p_1)$ [3, 7, *DET*]

Puisque la ligne 8 de cette preuve correspond à la première ligne de la dérivation de T5 donnée plus haut, nous disposons maintenant d'une preuve formelle de T5 dans L^{*Def}. Cet exemple illustre bien le caractère *interne* des définitions dans un tel système :

(1) A titre de thèses de L^{*Def}, les expressions définitoires des lignes 1 et 2 constituent des formules officielles du langage objet. Aussitôt la définition posée, le nouveau symbole # se trouve, de fait, intégré aux constantes du langage objet. L'ensemble des formules du système est alors élargi aux expressions de la forme $(\alpha \# \beta)$.

(2) Comme on peut le voir aux lignes 3-8, la façon dont la définition se trouve exploitée dans une preuve est entièrement réglée par les moyens officiellement déclarés dans la base axiomatique.

(3) Le fait que T5, qui n'est pas une thèse de L^*, soit prouvable dans L^{*Def} montre que les définitions jouent un rôle prépondérant dans les moyens déductifs de ce dernier système.

2. Créativité

La découverte par Leśniewski du caractère créatif des définitions explicites déclancha des réactions qui furent, dès les années vingt, l'objet d'une polémique au sein de l'Ecole de

Varsovie[11] : d'un côté Leśniewski adoptait des systèmes logiques incluant une règle définitoire et soutenait « que les définitions devaient être les plus créatives possibles », de l'autre Lukasiewicz rejetait les définitions créatives, soutenant qu'elles n'étaient rien moins que des « axiomes masqués » et qu'il fallait donc s'en tenir aux définitions purement abréviatives de Whitehead et Russell[12]. La formule T5 de notre exemple ci-dessus ne comportant aucune occurrence de terme défini, la définition de # est bel et bien *créative* au sens que Lukasiewicz et Leśniewski donnèrent à cet adjectif dans la littérature. En clair :

> Une définition *D* d'une constante *d* est **créative** dans un langage *S* si et seulement si il existe une formule *F* de *S*, sans occurrence de *d* et telle que *F* n'est prouvable dans *S* qu'avec l'aide de *D* (ou d'une autre définition).

Traditionnellement l'innocuité des définitions explicites était conçue comme allant de pair avec le caractère éliminable des définitions abréviatives. Or ce que les réflexions de Lesniewski permettent de montrer, c'est qu'une définition éliminable peut néanmoins être créative et donc jouer un rôle incontournable dans les moyens démonstratifs d'un système formel. Reste que pour justifier le recours aux définitions internes dans notre programme logiciste (Gessler, Joray, Degrange 2005), il faut également dégager avec précision *ce qu'*elles introduisent dans le formalisme, afin de déterminer s'il ne convient pas d'y voir, comme le pensait Lukasiewicz, des axiomes cachés.

2.1. L'argument de Lukasiewicz

Pour montrer que l'inscription de définitions créatives relève d'un ajout implicite d'axiomes, Lukasiewicz s'appuyait sur des

---

11 Sur les circonstances et l'objet de cette polémique, *cf.* Joray (2006).

12 L'échange que nous reprenons ici entre Lukasiewicz et Leśniewski se trouve consigné dans les comptes rendus de Lukasiewicz (1928).

exemples construits dont le plus caractéristique fut publié en 1939, dans son article sur le calcul équivalentiel[13]. Dans la section qu'il consacre à la définition créative, Lukasiewicz présente un système biconditionnel assez étrange, constitué d'une tautologie posée à titre d'unique axiome, ainsi que de trois règles d'inférence (le détachement du biconditionnel E, la substitution et une règle autorisant l'inscription de définitions sous forme d'expressions biconditionnelles). Dans son écriture préfixée, le système, que nous nommons ici Luk^{Def}, se présente ainsi :

Système Luk^{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*

Lukasiewicz montre alors que les seules thèses que l'on peut prouver sans usage de la règle *DEF* sont les instances de substitution de l'axiome (la forme particulière de celui-ci excluant en effet l'application de *DET*). Puisque de nombreuses tautologies biconditionnelles ne sont pas de telles instances de substitution, il apparaît que Luk^{Def} amputé de *DEF* serait un système *incomplet* du calcul biconditionnel classique. En particulier, les deux tautologies suivantes n'y seraient pas démontrables :

(*) $\mathrm{E}pp$

(**) $\mathrm{EE}pq\mathrm{EE}rq\mathrm{E}pr$

A noter que (**), connue comme « formule de Meredith », est suffisante, à titre d'axiome unique et associée à *DET* et *SUB*, pour obtenir un calcul biconditionnel complet.

S'autorisant désormais l'usage de *DEF* dans Luk^{Def}, Lukasiewicz montre alors par la preuve qui suit que (**), qui

13 Lukasiewicz (1939).

n'était pas prouvable sans *DEF*, le devient à la suite de la définition d'un connecteur unaire (le *verum*, noté ici V) :

1. EVpEpp [DEF]
2. **EEV*p*E*pp***EEV*p*E*pp*EE*pq*EE*rq*E*pr* [Ax, *SUB* (*s* / V*p*)]
3. **EEV*p*E*pp***EE*pq*EE*rq*E*pr* [1, 2, *DET*]
4. EE*pq*EE*rq*E*pr* [1, 3, *DET*] (Formule de Meredith)

D'incomplet qu'il était sans l'usage de *DEF*, le système Luk^{Def} devient donc complet par l'effet créatif de la définition posée en ligne 1. Pour Lukasiewicz, la définition revient ici à un ajout caché d'axiome car elle ne se limite pas à introduire le connecteur V ; elle participe implicitement, mais de manière cruciale, à la caractérisation du connecteur primitif E.

Si Lukasiewicz a raison de soutenir qu'une telle créativité doit être le fait d'un axiome déclaré, il se trompe cependant sur un point important qui aurait rendu son exemple de définition inacceptable aux yeux de Leśniewski : la définition de V n'est pas ici une définition *explicite.* En effet, dans le contexte où la définition se trouve inscrite, la tautologie (*) n'est pas encore dérivable et cela signifie que la définition posée en ligne 1 exprime, par le biais de E, une relation entre *definiendum* et *definiens* qui n'est pas une relation d'équivalence car elle n'est pas même réflexive. Si l'exemple de Lukasiewicz présente bien un cas de créativité nécessitant l'inscription de la définition à titre d'axiome, aucune conclusion ne peut en être tirée quant à la créativité des définitions explicites. En outre, nous avons montré (Joray 2005b) que dans le cadre de logiques propositionnelles simples (sans quantification ni variables de connecteurs) aucune définition explicite ne peut être créative.

2.2. Le potentiel de créativité

Face à la forte créativité de l'exemple de Lukasiewicz, celle des définitions explicites est beaucoup plus modeste. Tout d'abord, il convient de souligner qu'une définition explicite n'est pas susceptible d'introduire une signification qui aurait été jusque là absente du langage où elle se trouve posée. En effet, dans une définition explicite, c'est le *definiens* qui exprime la signification qui va être attachée au *definiendum*. Or, puisque le *definiens* ne doit contenir que des termes primitifs ou déjà définis, cette signification est forcément disponible avant l'inscription de la définition.

Ce qui se trouve modifié par contre, c'est le mode d'expression de cette signification. Dans la série des langages L, L^* et L^{*Def} discutés en section 1, il est possible, sans aucun recours à des définitions et par composition des connecteurs primitifs, d'exprimer l'application de n'importe quelle fonction de vérité à des propositions données. Ce que la définition introduit, c'est la possibilité de désigner ces fonctions par le biais d'une appellation qui leur est propre. Dès lors, plutôt que de simplement user de la fonction, il est désormais possible d'exprimer quelque chose à son sujet, c'est-à-dire de lui faire jouer le rôle d'argument.

Pour que cette nouveauté soit féconde, il convient cependant que le langage où se trouve inscrite la définition présente des moyens syntaxiques permettant de l'exploiter. Ainsi la définition de # reste-t-elle stérile dans le langage simple du système L, alors qu'elle est créative dans l'expansion L^* dont les moyens syntaxiques de base sont plus développés.

3. Abstraction et appellation

L'idée que les définitions explicites sont purement abréviatives et ne peuvent introduire, comme le disent les auteurs des *Principia*, que des « facilités typographiques » (1927, 11) est très réductrice. Poser une pure abréviation, c'est par exemple décider de nommer à fin de brièveté « CdRS » ce qui jusqu'alors était nommé « Centre de Recherches Sémiologiques », ou encore d'user du signe « °C » en lieu et place des mots « degré Celsiu s». On ne procède à une pure abstraction que lorsque l'on remplace par une nouvelle appellation (généralement plus courte) une appellation déjà existante et de la même catégorie syntaxique.

Les définitions explicites font, pour la plupart, beaucoup plus que de poser une abréviation. Dans notre définition de # :

$$\text{(Df\#)}\quad P \# Q =\text{df} \sim(P \supset Q)$$

le *definiens* « $\sim(P \supset Q)$ » ne contient aucune appellation de la fonction que nous voulons associer au nouveau symbole #. L'expression définitoire n'est qu'un résumé d'une procédure complexe dont l'abréviation n'est qu'une phase relativement secondaire sur le plan théorique. Cette procédure peut être décrite par les cinq étapes suivantes :

1. On sélectionne une expression du fait de l'intérêt que l'on entend porter à telle ou telle signification. Cette expression servira de *definiens*. Dans notre exemple on sélectionne

$$\sim(P \supset Q).$$

2. On procède à une analyse grammaticale de l'expression retenue. Les règles de formation du langage concerné doivent permettre de déterminer l'appartenance catégorielle de chacun de ses constituants, ainsi que leur statut respectif de constante ou de variable. En résumé, dans notre exemple, on doit disposer des paramètres présents dans le diagramme suivant :

$\sim$ (	P	$\supset$	Q	)	Catégorie de l'expression : s
s/s	s	s/ss	s		
$\mathrm{C}^{\mathrm{ste}}$	$\mathrm{V}^{\mathrm{ble}}$	$\mathrm{C}^{\mathrm{ste}}$	$\mathrm{V}^{\mathrm{ble}}$		

3. On *abstrait* les éléments variables et on obtient une expression fermée dont les paramètres dégagés lors de l'étape 2 permettent de déterminer la catégorie. Cette expression fermée constitue une *appellation* de la signification visée. Dans notre exemple, on abstrait le couple d'arguments $\langle P, Q\rangle$ et l'expression fermée obtenue est de catégorie *s/ss* (connecteur propositionnel binaire) :

   $\langle P, Q\rangle[\sim(P \supset Q)]$ Catégorie de l'appellation : *s/ss*

4. On impose un symbole comme *abréviation* de l'appellation obtenue à l'étape 3 par abstraction. L'abréviation reçoit la même catégorie que l'appellation qu'elle abrège Dans notre exemple :

   $\# := \langle P, Q\rangle[\sim(P \supset Q)]$

5. On *applique* le nouveau symbole aux mêmes arguments que ceux abstraits à l'étape 3 et ceci en conformité aux règles de formation relatives aux éléments de sa catégorie dans le langage concerné. On obtient alors le *definiendum*. Dans notre exemple, les règles imposent qu'un élément de catégorie *s/ss* reçoive ses arguments à gauche et à droite. On obtient donc le *definiendum* :

   $$P \# Q$$

Cette analyse montre que les deux mécanismes sous-jacents à l'inscription d'une définition explicite sont l'abstraction et l'imposition d'une pure abréviation. Mais comme l'imposition relève de l'étiquetage et non d'un processus de créativité, l'aspect novateur des définitions leśniewskiennes doit être imputé au mécanisme de l'abstraction. Plus précisément, en faisant de la définition explicite un outil interne aux formalismes, on augmente la puissance des systèmes axiomatiques en donnant accès aux possibilités qu'offre l'abstraction, alors que ces possibilités restent inopérantes avec les définitions externes.

Conclusion

Paradoxalement nous obtenons donc que l'opposition historique que nous avons retracée entre les positions théoriques de Lukasiewicz et Leśniewski se doit d'être nuancée. En effet, si nous ne pouvons concéder à Lukasiewicz l'idée qu'un réel impact des définitions explicites créatives les placerait au même statut que les axiomes, nous avons cependant montré que le potentiel des définitions créatives internes se limite aux effets d'une procédure d'abstraction. Comme cette procédure est de nature purement logique, nous avons ainsi montré que les systèmes de Leśniewski constituent une base adéquate en vue d'une construction logiciste de l'arithmétique.

Références bibliographiques

EBERT P. A. & ROSSBERG M. (2007). What is the Purpose of Neo-Logicism ? In : Joray (ed.), *Contemporary Perspectives on Logicism and the Foundation of Mathematics*. Université de Neuchâtel : Travaux de Logique 18, 33-61.

GESSLER N., JORAY P., DEGRANGE C. (2005). *Le logicisme catégoriel*. Université de Neuchâtel : Travaux de Logique 16

GESSLER N., MIÉVILLE D., PEETERS M. (2001-07). *Introduction à l'œuvre de S. Leśniewski*. Université de Neuchâtel : Travaux de Logique, cinq cahiers spéciaux.

HALE B. & WRIGHT C. (2001). *The Reason's Proper Study. Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford : Clarendon.

JORAY P. (2005a). Should Definitions be Internal ? In : M. Bilkova & L. Behounek (eds), *The Logica Yearbook 2004*. Praha : Filosofia, 189-99.

JORAY P. (2005b). *What is wrong with creative definitions ? Logica* (Univ. de Wroclaw), 23.

JORAY P. (2006). *La définition dans les systèmes logiques de Lukasiewicz, Leśniewski et Tarski*. In : Pouivet & Rebuschi 2006, 203-222.

JORAY P. (2007). A New Path to the Logicist Construction of Numbers. In : Joray (ed.), *Contemporary Perspectives on Logicism and the Foundation of Mathematics*.Université de Neuchâtel: Travaux de Logique 18, 147-65.

LEŚNIEWSKI S. (1931). Ueber Definitionen in der sogenannte Theorie der Deduktion. In: Leśniewski 1992, 629-648.

LEŚNIEWSKI S. (1992). *Collected Works*. Surma S. J., Srzednicki J. T. & Barnett D. I. (eds). Dordrecht : Kluwer / Warszawa : PWN.

LUKASIEWICZ J. (1928). Sur les définitions dans les systèmes déductif / Le rôle des définitions dans les systèmes déductifs (2 cptes rendus de conférences). [Trad. fr. de Blaszczyk M., en annexe à Joray 2006].

LUKASIEWICZ J. (1939). The Equivalential Calculus, In : Lukasiewicz, *Selected Works*, L. Borkowski (ed). Amsterdam : North Holland / Warszawa : PWN, 1970, 250-277.

MIÉVILLE D. (1984). *Un développement des systèmes logiques de Stanislaw Leśniewski. Protothétique – Ontologie – Méréologie.* Berne : Peter Lang.

POUIVET R. & REBUSCHI M. (éds) (2006). *La philosophie en Pologne 1918-1939.* Paris : Vrin.

RICKEY V. F. (1975). Creative Definitions in Propositional Calculi. *Notre-Dame Journal of Formal Logic 16*.2, 273-294.

TARSKI A. (1923). Sur le terme primitif de la logistique. In : *Logique, Sémantique, Métamathatique*, T.1. Paris : A. Colin. [Trad. fr. de Granger G.-G.].

WHITEHEAD A N. & RUSSELL B. (1927). *Principia Mathematica.* Cambridge : CUP (2nd edition; 1st ed. 1910).

WRIGHT C. (1983). *Frege's Conception of Numbers as Objects.* Aberdeen University Press.