

# L'argument de Fodor et l'opérateur de Hilbert\*

Jean Petitot  
CAMS, EHESS

Octobre 1975

## Résumé

Nous discutons un argument de Jerry Fodor sur les conditions de possibilité de l'apprentissage de la logique en utilisant l' $\varepsilon$ -calcul de David Hilbert qui est équivalent au calcul des prédicats.

J'aimerais revenir sur l'intervention de Jerry Fodor concernant l'acquisition de concepts et l'impossibilité d'acquérir par apprentissage des structures plus puissantes que celles dont on dispose au départ, et en particulier sur son argument logique central que Jacques Monod et Anthony Wilden ont déjà discuté.

## 1 L'argument de Fodor

Si je l'ai bien compris, l'argument est le suivant : si l'on considère les capacités logiques d'un organisme comme une succession de "stades" logiques et si cette succession est un ordre de réalisations de logiques de complexité et de capacité *strictement* croissantes, alors les transitions entre stades ne peuvent pas correspondre à un processus d'apprentissage et l'hypothèse innéiste est donc, en dernière instance, la seule acceptable.

Pour exemplifier cela, Fodor s'est référé à la suite  $S_1, S_2$  d'un stade  $S_1$  correspondant à une logique propositionnelle et d'un stade  $S_2$  correspondant à une logique des prédicats du premier ordre (donc avec quantificateurs), logique qui est effectivement d'une complexité syntaxique et d'une force déductive strictement supérieures. Il ne peut, selon lui, y avoir de processus d'apprentissage de  $S_1$  vers  $S_2$  car dans  $S_2$  il existe des formules quantifiées du type "pour tout  $x$ ,

---

\*Exposé au débat entre Jean Piaget et Noam Chomsky organisé par Massimo Piatelli-Palmarini au Centre Royaumont les 10-13 octobre 1975. Étaient présents, entre autres, Scott Atran, Gregory Bateson, Norbert Bischof, Guy Cellier, Jean-Pierre Changeux, Antoine Danchin, Jerry Fodor, Maurice Godelier, Bärbel Inhelder, François Jacob, André Leroi-Gourhan, Jacques Mehler, Jacques Monod, Seymour Papert, David Premack, Hilary Putnam, Dan Sperber, René Thom, Stephen Toulmin, Anthony Wilden, Thomas de Zengotita

$F(x)$ ” alors que l’on ne peut pas exprimer *dans*  $S_1$ , pour une telle formule, une condition de vérité du genre “pour tout  $x$ ,  $F(x)$  est vraie si et seulement si ...”. Ce point est le cœur de l’argument et c’est lui que j’aimerais discuter.

Pour ce faire, je vais considérer que  $S_1$  est le calcul des prédicats sans quantificateurs qui fonctionne comme un calcul propositionnel. L’arrière-fond logique de mes remarques est par conséquent le problème de *l’élimination des quantificateurs* dans les théories logiques du premier ordre.

Bien qu’il ne s’agisse pas dans notre débat vraiment de logique mais plutôt de psychologie, l’argument de Fodor repose cependant sur un résultat de logique élémentaire, résultat qui repose lui-même sur ce que j’appellerai une conception “standard” des théories logiques. Dans une certaine mesure, Fodor “absolutise” cette description standard et à partir de là invalide, de façon semble-t-il inattaquable, une théorie des stades. Mais pourtant il ne va pas de soi que cette conception standard soit “psychologically rooted”. Car il pourrait bien se faire qu’il existe une description, disons non standard, du stade  $S_2$  qui au lieu d’enrichir la syntaxe de  $S_1$  par la quantification enrichisse ses propriétés sémantiques (c’est-à-dire l’interprétation des termes intervenant dans les prédicats, fonctions, relations, etc.) d’une façon originale, ce qui pourrait permettre son apprentissage à partir de  $S_1$ .

Je rejoins ici la remarque d’Anthony Wilden, à savoir que  $S_2$  est peut-être syntaxiquement plus riche que  $S_1$ , mais n’en est pas un enrichissement sémantique.

Or, il se trouve qu’une telle description non standard existe. Elle est due à Hilbert et à ses disciples Paul Bernays et Wilhelm Ackermann.<sup>1</sup> Bien que Hilbert l’ait introduite pour des raisons sophistiquées de métamathématique, je ne pense pas qu’elle se borne à être une simple curiosité technique.<sup>2</sup> Bien au contraire.

## 2 Le calcul de Hilbert

Pour exposer en quelques mots la description hilbertienne, je rappelle d’abord l’opposition syntaxe/sémantique en théorie élémentaire des modèles.

(i) La syntaxe est la logique des prédicats du premier ordre avec des symboles de constantes, de fonctions et de relations caractéristiques (par exemple 0, 1, +,  $\times$ , pour l’ensemble des entiers  $\mathbb{N}$ ).

(ii) La sémantique, est un ensemble muni d’une structure (par exemple une structure d’anneau commutatif pour  $\mathbb{N}$ ) où ces symboles prédictifs et fonctionnels caractéristiques sont interprétés en termes des opérations définissant la structure.

(iii) Les formules quantifiées s’interprètent de façon standard.

---

<sup>1</sup>Cf. Hilbert-Bernays [9].

<sup>2</sup>Le théorème de Gödel ayant drastiquement redimensionné le rêve de Hilbert d’une fondation métamathématique finitiste des mathématiques, sa version du calcul des prédicats n’est plus guère utilisée. Pour des précisions sur le programme de Hilbert et sur le contexte, cf. Kreisel [10] et Van Heijenoort [7].

Hilbert introduit alors pour chaque prédicat (disons unaire<sup>3</sup>)  $F(x)$ , de façon *a priori* et purement syntaxique, un *individu*, qu'il appelle un  $\varepsilon$ -terme et qu'il note  $\varepsilon_x F(x)$ , où la variable  $x$  devient liée par l'opérateur  $\varepsilon$ . Et il interprète ce terme individuel en disant que  $\varepsilon_x F(x)$  représente un élément satisfaisant  $F$  mais au demeurant *quelconque*. On peut préciser cette interprétation de deux façons, *de re* et *de dicto* (pour reprendre une vieille opposition scholastique).

(1) Soit l'on considère l'opérateur  $\varepsilon$  *de re* comme un opérateur de *choix*, choisissant un certain élément de l'extension  $X_F$  du prédicat  $F$ .

(2) Soit l'on considère  $\varepsilon_x F(x)$  *de dicto* comme un élément *générique et typique* de  $X_F$ , comme une "idée in individuo" hypostasiant le *concept* de  $F$  dans un individu. *De dicto*,  $G(\varepsilon_x F(x))$  signifie que tous les éléments de  $X_F$  satisfont  $G$ , autrement dit que  $X_F \subset X_G$ .<sup>4</sup>

C'est l'interprétation *de dicto* que je privilégie ici car l'usage d'éléments *génériques* et *typiques* est massif dans la langue naturelle et la psychologie de l'acquisition des concepts. Il est donc essentiel de voir comment il peut être introduit en logique formelle et avec quelles conséquences.

Le premier intérêt de l'opérateur  $\varepsilon$  est d'introduire des identités syntaxiques qui ne sont ni des symboles d'individus au sens standard  $a, b, c$ , etc... ni des symboles de variables  $x, y, z$ , etc. Il introduit dans la symbolique logique la question linguistique si délicate des déterminants (*un* objet qui ..., *le* seul objet qui ...). Et cela de façon tout à fait non triviale. En effet, un  $\varepsilon$ -terme  $\varepsilon_x F(x)$  ne correspond pas forcément à un article défini ou à un article indéfini.

L'article défini correspond à l'opérateur  $\iota$  de Russell. Si  $F(x)$  est un prédicat caractérisant un individu, i.e. un prédicat pour lequel il y a existence et unicité de l'individu qui y satisfait (par exemple être un nombre entier pair et premier, ce qui caractérise le nombre 2), Russell note  $\iota_x F(x)$  cet individu.

Quant à l'article indéfini, il correspond à l'opérateur  $\eta$  de Hilbert. Étant donné un prédicat  $F$  pour lequel la formule  $\exists x F(x)$  est dérivable, on introduit le terme  $\eta_x F(x)$  représentant *un* des individus de  $X_F$  (qui est donc non vide). Ce terme satisfait évidemment par définition la formule  $F(\eta_x F(x))$ .

L'interprétation linguistique des  $\varepsilon$ -termes comme termes associés à des déterminants est alors la suivante. Pour un prédicat  $F$  *quelconque*, i. e. soumis à aucune restriction sur l'existence et l'unicité de ses "solutions", l'individu  $\varepsilon_x F(x)$  est par définition l'individu

$$\eta_x (\exists y (F(y) \rightarrow F(x))).$$

<sup>3</sup>Pour simplifier l'argumentation, je me restreindrai aux prédicats unaires. La complexité technique augmente évidemment considérablement avec des prédicats  $n$ -aires et emboîtés.

<sup>4</sup>En géométrie algébrique on sait définir rigoureusement des points génériques. Une variété algébrique (affine)  $V$  de  $K^n$  définie sur un corps de base  $K$  (commutatif et algébriquement clos) est définie par un idéal  $J$  de l'anneau de polynômes  $P = K[x_1, \dots, x_n]$  (celui des polynômes dont l'annulation définit  $V$ ). Cet idéal possède un "radical"  $\sqrt{J}$  qui est l'idéal des éléments  $p$  de  $P$  tels qu'une puissance  $p^n$  de  $p$  appartienne à  $J$ . Les points au sens classique de  $V$  correspondent aux idéaux *maximaux* de  $P$  contenant  $\sqrt{J}$  alors que les points génériques correspondent aux idéaux *premiers* contenant  $\sqrt{J}$ . Si  $V$  est irréductible,  $\sqrt{J}$  est un idéal premier et correspond au point générique de  $V$ .

Cette définition a bien un sens pour *tout* prédicat unaire, car la formule

$$\exists x (\exists y (F(y) \rightarrow F(x)))$$

est *toujours* dérivable.

Cette interprétation “linguistique” est au fond assez subtile car elle involue, dans l’écriture même du terme  $\varepsilon_x F(x)$  la question de son existence.

Hilbert introduit alors l’axiome dit  $\varepsilon$ -formula :

$$(\varepsilon\text{-formula}) \quad F(a) \rightarrow F(\varepsilon_x F(x))$$

et analyse le système formel ainsi obtenu où si  $F$  et  $G$  sont deux prédicats équivalents, les  $\varepsilon$ -termes  $\varepsilon_x F(x)$  et  $\varepsilon_x G(x)$  sont égaux. Dans ce système, appelé  $\varepsilon$ -calculus et qui est celui adopté par Bourbaki dans son premier tome *Théorie des ensembles* de ses *Éléments de mathématique*<sup>5</sup>, on définit le quantificateur existentiel par l’équivalence purement syntaxique

$$\exists x F(x) \stackrel{\text{def}}{=} F(\varepsilon_x F(x))$$

i.e. l’individu  $\varepsilon_x F(x)$  satisfait effectivement  $F$ . Le fait capital à souligner ici (pour la discussion de l’argument de Fodor) est que l’introduction des  $\varepsilon$ -termes permet d’exprimer comme propriétés d’individus (donc dans un calcul de type  $S_1$ ) des *formules quantifiées* de la logique du premier ordre. C’est précisément ce que voulait d’ailleurs Hilbert dans l’espoir de pouvoir contrôler par une métamathématique finitiste le caractère infinitiste des quantifications.

Mais si l’on applique alors l’équivalence

$$\forall x F(x) \leftrightarrow \neg \exists x (\neg F(x))$$

( $\neg$  étant le symbole de la négation) on voit que l’on a

$$\forall x F(x) \leftrightarrow F(\varepsilon_x (\neg F(x))) .$$

Dit naïvement, cela signifie que si l’individu quelconque satisfaisant non- $F$  satisfait “quand même”  $F$  alors *tous* les individus satisfont  $F$ . On remarquera que les arguments du type “si même lui alors tous” sont très courants dans la langue naturelle et sont psychologiquement spontanés et innombrables.<sup>6</sup>

Mais, pour en revenir à l’argument de Fodor, on voit que cette équivalence  $\forall x F(x) \leftrightarrow F(\varepsilon_x \neg F(x))$  fournit exactement l’expression dans  $S_1$  (i.e. en termes de propriétés d’individus) d’une condition de vérité d’une formule quantifiée universelle de  $S_2$ , à condition d’admettre dans  $S_1$  les individus génériques “non standard” que sont les  $\varepsilon$ -termes.

On voit donc la question. La logique standard du premier ordre et le calcul de Hilbert sont équivalents, ou plus précisément, le  $\varepsilon$ -calculus hilbertien est

<sup>5</sup>Cf. [2].

<sup>6</sup>L’exemple de Hilbert était : “Si même Aristide, le plus incorruptible des hommes, est corruptible, alors tous les hommes sont corruptibles”.

une ce que l'on appelle une "extension inessentielle" du calcul des prédicats. Il s'agit là d'un métathéorème non trivial. Mais en même temps, cet  $\varepsilon$ -calcul correspond à un système, disons  $S_1^\varepsilon$ , qui reste syntaxiquement de type  $S_1$ .

$S_1$  est donc effectivement strictement plus pauvre que  $S_2$  à condition d'interdire les  $\varepsilon$ -termes. On a ainsi le schéma suivant :

$$\begin{array}{ccc} S_1^\varepsilon & \xrightarrow{\cong} & S_2 \\ \uparrow \subset & & \nearrow \subset \\ S_1 & & \end{array}$$

Et je ne vois pas pourquoi l'on ne pourrait pas faire l'hypothèse que la transition  $S_1 \rightsquigarrow S_2$  est en fait une transition  $S_1 \rightsquigarrow S_1^\varepsilon$  suivie d'une transformation  $S_1^\varepsilon \rightsquigarrow S_2$  par appauvrissement sémantique des individus possibles et par enrichissement syntaxique corrélatif des types possibles de formules. Autrement dit, l'inclusion stricte de  $S_1$  dans  $S_2$  se *factoriserait* par  $S_1^\varepsilon$ .

Mais puisque  $S_1^\varepsilon$  est équivalent *formellement* à  $S_2$ , une telle hypothèse revient à poser que l'enrichissement syntaxique de  $S_1$  par la quantification syntaxique de  $S_1$  par la quantification syntaxique équivaut *formellement* à un certain enrichissement sémantique. Toutefois  $S_1^\varepsilon$  est très différent de  $S_2$  sémiotiquement. Dans notre contexte, la question devient alors de savoir si le stade  $S_1^\varepsilon$  ne serait pas, lui, dérivable de  $S_1$  par apprentissage. Dans ce cas, l'argument de Fodor s'appliquerait à  $S_2$  mais pas à  $S_1^\varepsilon$  qui lui est pourtant formellement équivalent.

Cela me conduit à la question de la *sémantique* de l'opérateur  $\varepsilon$ , qui soulève des problèmes fort intéressants.

### 3 La sémantique de l'opérateur $\varepsilon$

Le relatif désintéret dont a fait l'objet l'opérateur  $\varepsilon$  vient du fait qu'il n'existe pas de sémantique simple. Comme nous parlons ici de sémantique, introduisons donc un ensemble  $M$  (univers du discours) muni d'une structure à partir de laquelle les formules du calcul des prédicats associés sont interprétables. L'interprétation habituelle des  $\varepsilon$ -termes est *de re* et consiste à introduire ce que l'on appelle une *fonction de choix*, c'est-à-dire une fonction  $\Phi : \mathcal{P}(M) \rightarrow M$  de l'ensemble  $\mathcal{P}(M)$  des parties de  $M$  dans  $M$  associant à chaque partie  $X$  non vide de  $M$  un élément  $\Phi(X)$  qui appartient à  $X$  et représente l'élément choisi par  $\Phi$  dans  $X$ . Il faut noter que ces fonctions de choix peuvent varier suivant les "mondes possibles" et sont donc "intensionnelles".

Si alors le prédicat  $F$  est d'extension  $X_F$  non vide, i.e. si la formule  $\exists xF(x)$  est dérivable, il semblerait naturel d'interpréter l' $\varepsilon$ -terme  $\varepsilon_x F(x)$  comme ayant pour référent l'élément  $\Phi(X_F)$  de  $X_F$ . Mais dans ce cas on fait la sémantique de l'opérateur  $\eta$  et non pas celle de l'opérateur  $\varepsilon$ . Or, toute la force de l'opérateur  $\varepsilon$  provient de ce qu'il est défini sans restriction aucune, c'est-à-dire même pour les prédicats dont l'extension est *vide*. Et cela est même absolument nécessaire puisque c'est cela qui permet de définir la quantification universelle. Mais

si la formule  $\forall x F(x)$  est dérivable on se heurte à la contradiction apparente suivante : l' $\varepsilon$ -terme  $\varepsilon_x(\neg F(x))$  doit être un individu pour pouvoir satisfaire à  $F(\varepsilon_x(\neg F(x)))$  tout en étant un individu *sans référent* puisque justement il n'existe pas d'individu satisfaisant  $\neg F$ . On dit qu'un tel  $\varepsilon$ -terme est un terme zéro, un "null term".

## 4 L'infinitésimale leibnizienne comme $\varepsilon$ -terme zéro

Un exemple typique d' $\varepsilon$ -terme zéro est le symbole  $dx$  de l'infinitésimale chez Leibniz. Soit  $\mathbb{R}$  le corps totalement ordonné des nombres réels. Sa structure d'ordre est *archimédienne* : elle satisfait l'axiome selon lequel tout nombre (aussi grand que l'on veut) est atteignable par tout nombre  $\neq 0$  (aussi petit que l'on veut) à condition d'ajouter ce dernier à lui-même un assez grand nombre de fois. Autrement dit,  $\mathbb{R}$  satisfait l'énoncé :

$$(A) \quad \forall y \in \mathbb{R}^+ \forall x \in \mathbb{R}^+ \exists n \in \mathbb{N} (nx > y)$$

(où  $\mathbb{R}^+$  est l'ensemble des nombres réels strictement positifs et  $\mathbb{N}$  l'ensemble des entiers naturels).

En passant des grands nombres à leurs inverses, cet axiome dit qu'il n'existe pas d'infinitésimale dans  $\mathbb{R}$  : tout nombre non nul (positif) aussi petit que l'on veut est plus grand qu'un autre nombre non nul. Il n'existe donc pas de nombre (positif) qui soit non nul et plus petit que tous les nombres strictement positifs. Autrement dit,  $\mathbb{R}$  satisfait l'énoncé :

$$(I) \quad \forall y ((y \neq 0) \Rightarrow \exists r ((r > 0) \wedge (|y| > r)))$$

(où  $|y|$  est la valeur absolue de  $y$ ).

En utilisant le fait que la négation d'une implication  $A \rightarrow B$  est la conjonction  $A \wedge \neg B$ , il est facile de voir que le concept leibnizien d'infinitésimale recouvre très exactement l' $\varepsilon$ -terme zéro associé à l'universelle (I). (I) est de la forme  $\forall y G(y)$  et est donc équivalente à l'énoncé hilbertien  $G(\varepsilon_{\neg G})$ .  $\varepsilon_{\neg G}$  correspond à l'idée d'un nombre différent de 0 et dont la distance à 0 est inférieure à tout nombre réel strictement positif : c'est exactement le  $dx$  leibnizien. L'axiome d'Archimède est alors équivalent à l'énoncé  $G(\varepsilon_{\neg G})$  assurant que toute infinitésimale est soit nulle soit finie : il n'existe pas de référent numérique pour l'idée in individuo d'infinitésimale symbolisée par  $dx$ . Comme le disait Leibniz,  $dx$  est une *fiction*, mais une fiction "bien fondée", une fiction "fondée en réalité". Et c'est sur cette fiction que repose tout le calcul différentiel...

Pour pouvoir légitimer l'usage de la notation leibnizienne, il faut pouvoir satisfaire les contraintes suivantes.

(a) Il faut que les infinitésimales soient des entités que l'on puisse traiter comme des nombres, c'est-à-dire auxquelles on puisse appliquer les opérations fondamentales. Cela est nécessaire pour que des expressions comme  $x + dx$ ,  $dy/dx$ , etc. aient un sens.

(b) Étant donnée une fonction  $f$  d'argument  $x \in \mathbb{R}$ , il faut que l'on puisse étendre automatiquement et canoniquement  $f$  aux infinitésimales. Cela est par exemple nécessaire pour définir l'accroissement infinitésimal  $df$  de  $f$  au voisinage de  $x$  par  $df = f(x + dx) - f(x)$ .

(c) Le symbole  $dx$  n'étant pas référentiable dans  $\mathbb{R}$  ( $\varepsilon$ -terme zéro) il faut qu'il le devienne dans une extension  ${}^*\mathbb{R}$  de  $\mathbb{R}$  qui soit, dans un sens à préciser, *indiscernable* de  $\mathbb{R}$ .

En appliquant à l'analyse certains théorèmes fondamentaux de la théorie générale des modèles, *l'analyse non standard*<sup>7</sup> a montré qu'il est possible de satisfaire à ces trois contraintes et que cette "légitimation" des "fictions" leibniziennes simplifiait parfois de façon notable la formulation de l'analyse.

## 5 Éléments singuliers

On peut évidemment prolonger la fonction de choix  $\Phi$  à l'ensemble vide  $\emptyset$  en définissant  $\Phi(\emptyset)$  comme un élément quelconque de  $M$ . On peut même, comme le fait par exemple Hermes<sup>8</sup>, donner des arguments en faveur de l'identification de  $\Phi(\emptyset)$  à  $\Phi(M)$ , ce qui correspond à l'égalité apparemment paradoxale

$$\varepsilon_x(x \neq x) = \varepsilon_x(x = x) .$$

L'instanciation des termes zéro est syntaxiquement possible mais sémiotiquement contradictoire.

Je dirais que dans ce qui le caractérise le plus éminemment, l'opérateur  $\varepsilon$  induit une déhiscence entre le sens (sémiotique) et la dénotation (référentielle)<sup>9</sup>, déhiscence dialectique dont Hilbert a formalisé le fonctionnement syntaxique.

Toute instanciation d'un terme zéro par un élément de  $M$  met cet élément en "exclusion interne" avec  $M$  ou encore induit un splittage de son identité entre l'identité syntaxique qu'il symbolise et son identité sémantique contradictoire.

D'où l'alternative suivante. Soit s'en tenir à l'équivalence syntaxique entre  $S_1^\varepsilon$  et  $S_2$  et laisser de côté la question de la référence, soit au contraire s'interroger sur la fonction éventuellement pertinente de cette déhiscence de l'identité des  $\varepsilon$ -termes zéro. La question devient ainsi de savoir si l'on peut repérer une opérativité concrète correspondant à la thématization et à la formalisation hilbertienne ou encore si l'opérateur  $\varepsilon$  et les effets "d'exclusion interne" qui en procèdent renvoient à une "structure" au sens de Jean Piaget, c'est-à-dire à des processus précédant leur formalisation et leur thématization.

## 6 Quelques applications formelles

Je pense qu'un certain nombre de faits plaident en faveur de cette dernière hypothèse. J'en énumère ici quelques uns par ordre de pertinence logique

<sup>7</sup>Cf. Robinson [13].

<sup>8</sup>Cf. Hermes [6].

<sup>9</sup>Entre *Sinn* et *Bedeutung*, pour reprendre ces notions de Gottlob Frege devenues classiques. Cf. Frege [5].

décroissante et d'intérêt anthropologique croissant.

(1) L'intérêt principal de l'opérateur  $\varepsilon$  est son indétermination. Celle-ci correspond au degré d'indétermination requis pour une description adéquate de l'*abstraction*. Un bon exemple est celui du concept de *cardinal*. Un cardinal doit en effet se concevoir à la fois comme un ensemble et comme une classe d'équivalence pour la relation d'équipotence définie sur la classe de tous les ensembles. Ces deux requisits entrent en conflit dans une conception standard. Si  $X$  est un ensemble, la seule chose que l'on puisse dire de son cardinal  $|X|$  est que  $|X|$  est équipotent à  $X$  et est égal au cardinal de tout ensemble équipotent à  $X$ . Une définition tout à fait naturelle possédant exactement ce degré d'indétermination est alors la suivante (où  $\sim$  note l'équipotence) :

$$|X| = \varepsilon_Y (Y \sim X) .$$

C'est la définition qu'a adoptée Rudolf Caranap dans [3] et il semble tout à fait légitime de faire l'hypothèse que cette description décrit en fait le processus même d'acquisition psychologique du concept de cardinal.

(2) Autre exemple. Soit  $F(x)$  une formule dans la théorie d'une structure  $M$ . On dit que  $F$  est "collectivisante" s'il existe un ensemble  $X$  dont les éléments sont les individus qui satisfont à  $F$ . En d'autres termes,  $F$  est collectivisante si la formule existentielle

$$\exists X \forall x (x \in X \leftrightarrow F(x))$$

exprimant que  $F$  possède une extension est dérivable. Si l'on introduit alors l' $\varepsilon$ -terme (du second ordre)

$$\widehat{F} = \varepsilon_X (\forall x (x \in X \leftrightarrow F(x))) ,$$

(c'est-à-dire l'idée de l'extension de  $F$ ) dire que  $F$  est collectivisante c'est dire que la formule

$$\forall x (x \in \widehat{F} \leftrightarrow F(x))$$

est dérivable, ce qui définit conceptuellement (abstraitement) l'extension de  $F$ .

(3) Un autre exemple important est celui de l'arithmétique. C'est d'ailleurs, répétons-le, pour démontrer dans le cadre de sa stratégie métalogue finie la consistance de l'arithmétique que Hilbert a introduit l'opérateur  $\varepsilon$ . Dans cette théorie, l'opérateur  $\varepsilon$  peut s'interpréter sémantiquement comme un "least number operator" : la fonction de choix  $\Phi$  associée à tout sous-ensemble d'entiers non vide  $X$  de l'ensemble des entiers  $\mathbb{N}$  son plus petit élément. Si l'on applique la règle d'Hermès  $\Phi(\emptyset) = \Phi(M)$  on a alors

$$\Phi(\emptyset) = \Phi(\mathbb{N}) = 0 .$$

Cela est intéressant dans la mesure où cela permet de faire le lien entre la théorie formelle de l'arithmétique et l'élaboration de Frege, si injustement critiquée par Russell, d'une déduction conceptuelle de l'arithmétique. En effet, en tant que  $\Phi(\emptyset)$ , le nombre zéro représente l' $\varepsilon$ -terme zéro  $\varepsilon_x (x \neq x)$  qui est justement le zéro au sens de Frege comme concept de l'objet non identique à lui-même.



## 7 Quelques applications plus philosophiques

Mais en fait je pense que l'intérêt de l'opérateur  $\varepsilon$  va au-delà des exemples formels précédents.

(4) Revenons en effet à ce qui fait obstruction à une interprétation univoque et consistante des  $\varepsilon$ -termes.. En détournant des expressions géométriques, je dirai que dans le cas d'un jugement universel  $\forall xF(x)$ , l' $\varepsilon$ -terme zéro  $\varepsilon_x(\neg F(x))$  qui est en situation "d'exclusion interne" est un élément *singulier* et *limite* tandis que l'élément  $\varepsilon_x(F(x))$  est un élément générique et typique. Le fait que ces deux éléments puissent être représentables par le même élément  $\Phi(M) = \Phi(\emptyset)$  manifeste une tendance à leur identification par "coincidentia oppositorum". Il s'agit là d'un phénomène sans doute assez profond relevant d'une sémiotique "dialectique" du concept et possédant peut-être la structure d'un "double bind" au sens de Gregory Bateson. On pourrait aussi évoquer un lien avec la logique "paraconsistante"<sup>10</sup> et tout un ensemble de logiques qui cherchent à formaliser des expressions linguistiques naturelles et spontanées qui ne sont pas formalisables en logique classique. Nous avons cité l'exemple de Hilbert "si même  $a$  alors tous" qui vient du fait que l'on identifie  $a$  au référent d'un terme zéro  $\varepsilon_{\neg F}$  et que l'on infère de  $F(a)$  à  $\forall xF(x)$ . On peut aussi citer les généralisations abusives partant du cas particulier  $F(a) \wedge G(a)$  pour inférer  $(\forall x \in X_F) G(x)$  qui viennent de l'identification de  $a$  à l' $\varepsilon$ -terme  $\varepsilon_F$ . De façon générale, c'est un processus psychique universel que d'idéaliser un individu  $a$  possédant une propriété  $F$  en un individu hypostasiant et incarnant idéalement  $F$ . Dans ce genre d'idéalisation, la généralité de l' $\varepsilon$ -terme  $\varepsilon_F$  est pensé intensivement comme la maximisation de  $F : a$  est "le plus"  $F$  qu'il est possible (cf. encore l'exemple de Hilbert sur Aristide).

(5) L'hypothèse d'une réalité psychique d'un double bind entre éléments singuliers-limites et éléments génériques-typiques reprend et précise l'hypothèse du neurologue et analyste hongrois Imre Hermann qui s'est proposé de hisser au rang de véritable fonction psychique les raisonnements à la Hilbert.<sup>11</sup> L'utilisation d'éléments singuliers paradoxaux pour justifier des équivalences plus ou moins rigoureuses du type  $\forall xF(x) \leftrightarrow F(\varepsilon_x(\neg F(x)))$  pululent en effet dans le discours psychotique. Dans son article sur la psychologie et la psychopathologie de la fonction psychique du choix, Hermann, en critiquant d'ailleurs au passage Vigotski, montre comment les raisonnements universalisant à partir d'éléments singuliers (qu'il appelle les "bons exemples") constituent une forme de pensée archaïque encore psychiquement très prégnante. Il va même jusqu'à les mettre en parallèle avec le concept pavlovien de phase ultra-paradoxa et jusqu'à caractériser la paranoïa par la prévalence de tels processus psychiques.

(6) En s'éloignant encore plus de la logique et de sa norme pour nous tourner vers l'anthropologie, je pense que l'on peut soutenir que la plupart des mythes dits de fondation ne font que *narrativiser* un fond structural primaire qui est celui de l'exclusion interne (par rapport au groupe social) d'un individu marqué, thématé comme héros fondateur. Cette narrativisation a pour fonction logique

---

<sup>10</sup>Cf. da Costa [4].

<sup>11</sup>Cf. Hermann [8].

concrète (au sens de la “logique concrète” des mythes mise à jour par Claude Lévi-Strauss<sup>12</sup>) de résoudre la question critique que rencontre tout groupe social, à savoir comment la relation d’appartenance sociale peut-elle être collectivisante.

## 8 Quelques conséquences

Quelles conséquences tirer de ces quelques remarques pour une théorie de l’apprentissage ?

Si l’on admet la conception hilbertienne, l’argument de Fodor n’est plus si directement évident et la conclusion innéiste qui en dérive devrait être reconsidérée. En effet, il ne s’agirait plus du passage du stade  $S_1$  au stade  $S_2$  mais d’un élargissement de  $S_1$  en  $S_1^\varepsilon$  par des termes génériques et des termes zéro, élargissement suivi d’une réorganisation de  $S_1^\varepsilon$  en  $S_2$  d’abord par élimination de ces termes génériques et de ces termes zéro (ce qui est un appauvrissement sémantique et pragmatique) puis par introduction des quantificateurs (ce qui est un enrichissement syntaxique). Mais si les éléments tant génériques que singuliers sont psychologiquement attestés, on peut se demander si l’élargissement de  $S_1$  en  $S_1^\varepsilon$  et la réorganisation de  $S_1^\varepsilon$  en  $S_2$  ne sont pas tous deux accessibles par apprentissage. Peut-être pas par un apprentissage au sens de Fodor mais peut-être par un apprentissage “en U” au sens de Mehler composé d’une perte suivi d’une acquisition.

## 9 Deux remarques finales

Deux points pour finir.

(1) Il me semble que l’analyse de l’opérateur  $\varepsilon$  comme structure au sens de Piaget peut éclairer quelque peu la question soulevée par Fodor de l’acquisition des concepts. Il faudrait ici revenir sur la façon dont Carnap a utilisé cet opérateur pour résoudre certaines difficultés de théories empiriques.

(2) Mon analyse de l’argument de Fodor n’est pas une critique de l’innéisme. La question posée par l’innéisme est celle de la façon dont on y fait recours. Il y a sans doute un noyau proto-logique élémentaire universel inné faisant partie de notre “initial state” cognitif. Mais pour ce noyau le qualificatif de “logique” est sans doute trop fort. En tenant compte de différentes formulations du calcul des prédicats on peut rendre vraisemblable dans la formulation  $S_1^\varepsilon$  un processus d’apprentissage qui semble invraisemblable dans la formulation  $S_2$ .

## 10 Conclusion

J’ai insisté sur le fait que le calcul de Hilbert est au niveau syntaxique une extension inessentielle du calcul des prédicats du premier ordre (métathéorèmes d’élimination du symbole  $\varepsilon$ ). On peut donc s’en passer au niveau de la logique

---

<sup>12</sup>Cf. Lévi-Strauss, [12].

formelle. Mais cela ne l'empêche pas d'être très intéressant sur le plan cognitif. Son interprétation sémantique (au sens de la théorie des modèles) est très riche et elle est en plus très pertinente et inspirante au niveau psycho-sémiotique.

Dans la logique des prédicats les quantificateurs sont des symboles primitifs alors qu'il est évident que la quantification n'est pas un phénomène psychologique primitif. L'intérêt de l'opérateur  $\varepsilon$  est de permettre de la dériver de fonctions de choix et d'éléments génériques typiques. Or on peut admettre que les processus de choix et de typicalisation appartiennent aux structures cognitives du sujet. La fonction psychique du choix appartient à la régulation sémiotique du concept.

C'est par normalisation de  $S_1^\varepsilon$  que la transformation  $S_1 \rightarrow S_2$  devient possible. Une normalisation n'est peut-être pas un apprentissage, mais ce n'est pas non plus forcément un processus programmé génétiquement.

## References

- [1] G. ASSER, Theorie der logischen Auswahlfunktionen, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 3,30-68, 1957.
- [2] N. BOURBAKI, Théorie des ensembles, *Éléments de mathématique, I*, Hermann, Paris, 1939/1954.
- [3] R. CARNAP, On the use of Hilbert's  $\varepsilon$ -operator in scientific theories, *Essays on the Foundations of Mathematics*, North-Holland, Amsterdam, 1962.
- [4] N. DA COSTA, On the Theory of Inconsistent Formal Systems, *Notre Dame Journal of Formal Logic*,15(4), 497-510, 1974.
- [5] G. FREGE, Über Sinn und Bedeutung, *Zeitschrift für Philosophie und philosophische Kritik*, 100, 22-50, 1892.
- [6] H. HERMES, *Eine Termlogik mit Auswunaloperator*, Lecture Notes n°6, Springer, Berlin,1965.
- [7] J. VAN HEIJENOORT (ed.), *From Frege to Gödel*, Harvard University Press, Cambridge, Mass.,1967.
- [8] I. HERMANN, *Psychoanalyse und Logik*, Internationaler Psychoanalytischer Verlag, Wien, 1924.
- [9] D.HILBERT, P.BERNAYS, *Grundlagen der Mathematik*, Vol.II, Springer, Berlin, 1939.
- [10] G. KREISEL, Hilbert's Programme, *Philosophy of Mathematics*, P. Benacerraf and H. Putnam eds., Prentice-Hall, Upper Saddle River, 1964.
- [11] A.C. LEISENRING, *Mathematical Logic and Hilbert's  $\varepsilon$ -symbol*, Mac Donald, London, 1969.

- [12] C. LÉVI-STRAUSS, *La Pensée sauvage*, Plon, Paris, 1962.
- [13] A. ROBINSON, *Non-standard Analysis*, Princeton University Press, 1974.