IA et Systèmes Formels

Pour représenter et résoudre une classe très générale de problèmes d'IA, le professeur Alain Mille vous a présenté (voir cours et TDs) un modèle général basé sur les notions d'espace d'états, d'opérateur permettant de passer d'un état à un autre, et de recherche d'une solution par exploration d'une partie de l'ensemble d'états.

La théorie des **systèmes formels** constitue un autre **cadre général** dans lequel on peut exprimer la plupart des problèmes d'IA.

Points à traiter :

- Définition de système formel et un exemple d'école
- Les interprétations (métasystèmes)
- Les limites internes du formalisme (le théorème de Gödel, 1931)
- Un système formel particulier : la Logique des Prédicats (rappels)

La notion de système formel est un outil essentiel que nous allons aborder ici.

• Un **système formel** (SF) est défini par 4 éléments :

- un **alphabet** : ensemble infini dénombrable de symboles (avec lesquels on peut construire des mots, des assemblages)
- un **procédé** : permettant de construire de suites **finies** de symboles (expressions, formules) qui seront acceptables ; alphabet + procédé = **langage**
- une **liste d'axiomes** (ou expressions de départ)
- des **règles de dérivation** (ou de déduction, ou d'inférence) pour déduire de nouvelles formules acceptables à partir d'autres formules déjà acceptées (on parle de théorèmes)

Question : qu'est-ce qu'un SF sait faire ? Quelles sont les tâches qu'il est sensé savoir réaliser ? **Réponse :** les SF ont une double tâche :

- déduire de nouvelles assertions (on parlera de théorèmes), créer de nouveaux théorèmes,
- dire si une expression donnée peut ou non se déduire des axiomes, i.e. dire s'il s'agit d'un théorème ou d'un non-théorème.

La plupart du temps, on utilise un SF **en l'interrogeant**, en lui posant des questions : est-ce que telle expression peut ou non se déduire des expressions de départ ?

Exemple. Nous allons commencer par un exemple très simple de SF. On considère :

- l'ensemble $\{a, b, \circ\}$ appelé l'alphabet
- l'ensemble des suites finies (la concaténation) de symbole(s) a, ou b ou $\circ$
- un axiome unique : $a \circ a$
- une règle unique : $C \circ D \rightarrow bC \circ bD$

Par convention, dans cette règle les symboles C et D sont des métasymboles pour abréger, ils désignent des suites quelconques de symboles a ou b.

Regardons ce système de **l'intérieur**. Pour engendrer les théorèmes il suffit de considérer de façon exhaustive toutes les applications possibles de la règle unique à partir de l'axiome unique. On obtient les preuves :

$$a \circ a$$
$$ba \circ ba$$
$$bba \circ bba$$
$$bbba \circ bbba$$
et ainsi de suite

Question : peut-on déduire le mot $ba \circ bba$? Le système ne donne pas de réponse.

Regardons ce système de **l'extérieur**. On voit que les théorèmes de ce SF coïncident exactement avec les mots de la forme : $b \ldots ba \circ b \ldots ba$, que nous pouvons abréger en : $b^p a \circ b^p a$

Par ailleurs, de façon évidente, $baab \circ abba$ ne saurait être un théorème.

• Les interprétations et métasystèmes (donner un sens)

Interpréter veut dire **donner un sens** aux différentes parties d'un SF. Autrement dit, sortir à l'extérieur du SF avec l'espoir d'obtenir plus d'informations. Interpréter un SF, c'est généralement le mettre en correspondance avec une réalité extérieure au système. On parle de **métasystème**.

Une **interprétation** est une correspondance I entre le système formel et un métasystème connu. Dans notre exemple, on peut prendre comme métasystème un sous-système de l'arithmétique usuelle.

On définit l'application I de la façon suivante :

I : a $\rightarrow$ 0 (le numéro zéro)
b $\rightarrow$ successeur d'un nombre
$\circ \rightarrow =$ (le signe égal)

On déduit que l'axiome s'interprète alors comme $0 = 0$, énoncé que nous considérons **vrai**. Les théorèmes successifs s'écrivent : $1 = 1, 2 = 2, \ldots, p = p$ et sont tous **vrais**.

Dans cet exemple on a aussi la réciproque, autrement dit une synchronisation parfaite entre vérité et théorème. Un énoncé comme $1 = 2$, qui saurait être l'interprétation de $ba \circ bba$ est pour nous, dans cette interprétation, visiblement faux.

Ceci veut dire que, dans cet exemple d'école, l'interprétation ≪colle≫ au système formel donné. Mais est-ce toujours le cas ? La réponse est **NON**.

• Les limites internes du formalisme. Le théorème de Gödel, 1931

L'arithmétique prend comme modèle l'ensemble des nombres entiers, que l'on note **N**.

On note **AF** un système formel de l'Arithmétique (ex: axiomes de Peano). Une année après avoir fini sa thèse, Gödel a montré qu'il n'existe pas une synchronisation parfaite entre le système **AF** et l'ensemble des vérités sur **N**. Autrement dit, **l'Arithmétique n'est pas complètement formalisable**, i.e. il existe des énoncés vrais dans l'interprétation standard (dans **N**) du langage de l'Arithmétique qui ne sont pas démontrables dans la théorie de Peano. Ce qui est vrai n'est pas toujours démontrable.

La preuve de ce résultat ne laisse aucun espoir. Si l'on décide de changer les axiomes de **AF** la démonstration marchera encore. Si l'on ajoute des axiomes (on peut penser que peut-être on n'a pas pris assez) la démonstration marchera encore. La seule solution serait d'ajouter tellement d'axiomes que le système deviendrait sans aucun intérêt (pratiquement toutes les expressions seraient des axiomes) et on perdrait la vérification mécanique des preuves.