

A LA RECHERCHE D'UNE GRAMMAIRE UNIVERSELLE

Gérolde STAHL

ABSTRACT

Since antiquity many philosophers and grammarians were looking for what is "behind" the particular grammars, for something like "the unchangeable principles common to all languages". Even limiting ourselves to the most concrete aspects of such a general grammar, we may ask whether there is something realizable among the risky hypotheses and the vague projects.

In this paper we do not try to discover something more or less hidden in the particular grammars, but to show, in a very general way, some directions for constructing, eventually, an universal grammar. Four approaches are mentioned:

- (1) artificial universal languages like Esperanto,
- (2) systems of automatic analysis of a language,
- (3) programming languages,
- (4) the first-order systems of logic.

It is shown how those approaches (and the experiences acquired in working with them) might be combined; but whether this combination produces interesting results and brings us nearer to an universal and rational grammar of our computerized epoch is an open question, which can be answered only by practical experience.

1. Introduction

Dès l'Antiquité beaucoup de philosophes et grammairiens ont cherché ce qu'il y a "derrière" les grammaires particulières, quelque chose comme "les principes immuables qui sont communs à toutes les langues"; cette grammaire générale d'une langue idéale pouvait être une entité métaphysique, un phénomène purement linguistique ou un système logique. Même si on se limite aux aspects les plus concrets d'une telle grammaire, il faut se demander s'il y a quelque chose de réalisable parmi les hypothèses risquées et les projets vagues. Sait-on avec plus ou moins de

précision ce qu'on veut avoir, à quoi on veut aboutir?

Il y a un siècle beaucoup de chercheurs auraient haussé les épaules et traité la question d'une "grammaire universelle" comme spéculation gratuite. Aujourd'hui, en considération de ce qui se passe en linguistique, informatique et logique actuelles, ils sont probablement moins critiques, à condition qu'on leur indique les intentions et les limites de la recherche.

Dans cet article on ne veut *découvrir* quelque chose qui se trouve plus ou moins caché dans les grammaires particulières, (ou discuter sur l'acquisition ou l'innéité des langues) mais plutôt montrer, d'une façon très générale, un mode de construction pour une éventuelle grammaire universelle, qui serait le produit d'une coopération de plusieurs disciplines et qui réunirait le maximum d'avantages qu'on peut attendre d'une telle coopération. Cette grammaire se limitera au langage écrit, et on parlera seulement de morphologie et syntaxe¹. On n'aura pas en vue un vocabulaire déterminé; mais, pour une grammaire idéalisée, le vocabulaire doit être soumis à certaines contraintes. Finalement on ne traitera pas les possibles applications pratiques.

Dans l'état du développement on distinguera, peut-être, quatre voies d'approximation qui pourraient aboutir un jour conjointement à une grammaire universelle.

2. Première voie, l'espéranto, etc.

A partir des langues naturelles existantes on crée une langue universelle artificielle qui a la même structure que les langues naturelles; l'espéranto est un bon exemple. La grammaire d'une langue idéale de ce type devrait avoir au moins les avantages suivants sur les grammaires des langues naturelles:

- (a) pas d'irrégularités morphologiques (par exemple, pas de verbes irréguliers);
- (b) pas d'ambiguïtés syntaxiques (comme "montre", qui peut être verbe ou substantif);
- (c) au moins certaines catégories d'expressions peuvent être reconnues immédiatement (par exemple, on reconnaît, sans risque d'erreur, les substantifs par leur terminaison).

Les points (a) et (b) sont visiblement absents des langues naturelles. En ce qui concerne (c), on en trouve très peu d'indicateurs plus ou moins

sûrs de catégories, même en des langues aussi "logiques" que le latin. La majuscule pour les substantifs allemands fait une heureuse exception; mais cet avantage est partiellement annulé au commencement d'une phrase (tous les mots avec majuscule) et dans certains autres cas.

L'espéranto n'a pas d'irrégularités morphologiques et probablement pas d'ambiguïtés syntaxiques (du moins je n'en ai pas trouvées). Il fait aussi un effort sérieux par rapport au point (c). Ainsi les substantifs ont la terminaison "o" (pluriel "oj", accusatif "on", "ojn"), les adjectifs ont la terminaison "a" (pluriel "aj", accusatif "an", "ajn"), les adverbes ont la terminaison "e" (mouvement dirigé "en") et les verbes à l'infinitif ont "i" (temps et modes "is", "as", "os", "us", "u"). Cependant l'espéranto n'est pas toujours aussi conséquent et a assez d'exceptions parmi les terminaisons. Ainsi certains prénoms féminins ont la terminaison "a", on a "tro" ("trop"), "tri" ("trois"), "kaj" ("et") et beaucoup d'autres cas qui empêchent l'exclusivité des terminaisons indiquées. Mais il serait injuste de ne pas reconnaître qu'un grand pas en avant a été réalisé.

3. Deuxième voie, les systèmes d'analyse automatique d'une langue

En relation avec la traduction assistée par ordinateur, on a créé, pour diverses langues, des systèmes d'analyse morphologique et syntaxique automatisée. Ces systèmes ont recours à un certain nombre de techniques grammaticales, comme par exemple à l'utilisation d'arborescences, et présentent, à la fin de l'analyse d'une langue, des structures graphiques très détaillées, où chaque sommet porte une ou plusieurs étiquettes. Des langues similaires peuvent aboutir à des structures très similaires, tandis que, pour des langues plus ou moins éloignées, on arrivera à des structures assez différentes.

Déjà les expériences acquises dans l'analyse morphologique et syntaxique de n'importe quelle langue peuvent être utiles dans l'approximation d'une grammaire universelle. Plus loin encore vont les essais de construction de structures intermédiaires communes à toutes les langues (quelquefois on parle de "langue intermédiaire" ou "langue pivot") ou plutôt communes à un grand nombre de langues. Quoique cette construction ne puisse pas être le but *pratique* de la traduction assistée par ordinateur, parce qu'un détour de ce genre serait très antiéconomique dans l'état actuel de l'informatique (voir [10]), elle a un intérêt *théorique* indéniable.

4. Troisième voie, les langues de programmation

Avec le développement des logiciels on s'est éloigné de plus en plus du langage binaire qui est propre au fonctionnement spécifique d'un type donné d'ordinateurs. Afin de faciliter l'accès à l'informatique, on a créé des langues de programmation assez avancées, comme FORTRAN, BASIC, etc., qui contiennent dans leurs instructions des mots d'une langue naturelle (de l'anglais, en l'occurrence).

Malgré leur apparence rudimentaire ces langues ont un pouvoir expressif considérable et constituent une caractéristique commune de beaucoup d'ordinateurs (elles sont des espèces d'espéranto pour les ordinateurs); en même temps elles établissent un premier pont entre langues d'ordinateur et langues naturelles. De cette façon les langues de programmation contribuent également à la constitution d'une grammaire universelle. Cela sera bien visible, quand, en section 6, on traitera la langue PROLOG.

5. Quatrième voie, les systèmes logiques de premier ordre

Les expressions bien formées des systèmes logiques de premier ordre (avec identité) et avec des constantes individuelles et fonctionnelles constituent aussi un type de langue qui est très simple et en même temps extraordinairement riche.

Un des avantages de cette langue est sa flexibilité; on peut présenter une phrase plus analysée ou moins analysée, selon les besoins du moment. Ainsi "Le père de Julie achète un livre à Metz" peut être simplement " $A\rho$ " (avec " A " unaire et " ρ " pour une constante individuelle) ou, avec beaucoup de détails, " $(\text{Ex})(\text{Lx} \wedge \text{Ap}(j), \underline{x}, \underline{m})$ " (avec " A " ternaire et " ρ " pour une fonction).

Les systèmes d'ordre supérieur (jusqu'à l'ordre ω) ne sont plus riches qu'en apparence, parce que tous ces systèmes peuvent être représentés dans un système de premier ordre à l'aide de prédicats spéciaux.

Contrairement à la mode actuelle de construire un système logique particulier pour tout problème, aussi mince qu'il soit, la plupart des problèmes développés dans ces systèmes peuvent être traités, et généralement d'une façon plus simple, dans les systèmes de premier ordre. Ainsi pour un grand nombre de problèmes intensionnels et épistémiques on élargit simplement l'univers du discours de premier ordre, qui de toutes façon est conventionnel (voir [4] et [7]). Pour beaucoup de

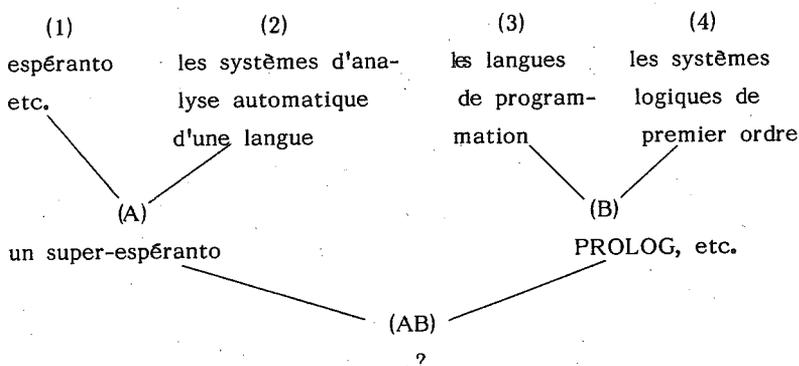
A LA RECHERCHE D'UNE GRAMMAIRE UNIVERSELLE

problèmes temporels et modaux on utilise simplement des prédicats de premier ordre plus complexes (voir [3],[5],[6], [8], et [9])². Cependant il y a des limites: par exemple, les problèmes qui ont recours à un niveau supérieur de langue ne peuvent pas être traités dans un système de premier ordre de niveau inférieur et exigent quelque chose de plus complexe.

Même en reconnaissant les limites de la logique de premier ordre, la possibilité de représenter, de façon tout à fait rationnelle, la plupart des situations (qui peuvent correspondre à des modèles très complexes), constitue aussi une voie qui mène vers une grammaire universelle.

6. Combinaison des quatre voies

Les quatre voies indiquées peuvent être combinées et ces combinaisons (dans le croquis qui suit, c'est (A), (B) et (AB)) nous rapprochent encore plus d'une grammaire universelle.



La combinaison (A) serait un développement de l'espéranto. Elle modifie l'espéranto actuellement existant et utilise pour cela les expériences acquises avec les systèmes d'analyse automatique, qui nous enseignent, par exemple, beaucoup sur les *constructions* ambiguës. En s'appuyant sur ces expériences, on arrivera à un degré d'analysabilité qui fera rêver un spécialiste en analyse automatique.

Pour cela il faut, parmi d'autres choses, une application stricte du point (c) de la section 2. En plus, on a besoin de ce qu'on pourrait appeler une "désambiguïsation des constructions": Par exemple, certaines expressions comme "de la patro" ("du père") peuvent être, comme en

français, un complément de nom ("le livre du père") ou un complément d'adjectif ("content du père") ou un complément du verbe ("venir du père", "être précédé du père"); un complément de nom ne dépend pas nécessairement du groupe nominal immédiatement précédent ("la maison de campagne du père", "la maison de la côte ouest de mon père"). Des solutions simples pour ce genre de problèmes existent, soit purement graphiques (en utilisant des parenthèses et d'autres signes), soit par modification du vocabulaire (par exemple, à la place du "de" on pourrait utiliser "ad", "ed", "id", "od", "ud", pour le complément d'adjectif, le complément de verbe etc.).

Ce "super-espéranto" serait un point de départ idéal pour des traductions presque totalement automatiques et très simples, vers un grand nombre de langues naturelles.

Dans la combinaison se rencontrent les langues de programmation et la logique de premier ordre. Une langue de programmation qui constitue un premier essai d'utiliser une partie des systèmes logiques mentionnés est PROLOG (programming in logic).³

Les phrases atomiques de la logique (avec symboles de constantes) comme " $a \text{ } \bar{F} \text{ } b$ " ou " $\bar{F}a \text{ } \& \text{ } c$ " se retrouvent en PROLOG exactement de la même façon comme " $a \text{ } \bar{F} \text{ } b$ " ("Pierre est-le-père-de Pauline") et comme " $\bar{F}(a \text{ } \& \text{ } c)$ " ("avait-donné (Pierre Pauline le-livre-jaune)").

Dans les questions posées en PROLOG comme "is (...)" et "which (x...)" ou dans certaines règles on trouve aussi des connecteurs comme "&", "not", "either... or" ou "if", tandis que les "x", "y", etc. qui ne suivent pas directement le "which(" correspondent pratiquement à des variables quantifiées existentiellement sur l'univers du discours du modèle constitué par les phrases atomiques entrées dans l'ordinateur. Le quantificateur universel, quand il précède une implication comme " $\bar{F}x = Gx$ " est représenté en PROLOG par "forall ".

PROLOG est seulement une première étape. On peut aller sûrement plus loin, si l'on veut combiner la logique de premier ordre avec une langue de programmation.

En ce qui concerne la combinaison entre (A) et (B), elle est sans doute réalisable. Donnera-t-elle des résultats intéressants? Nous rapprochera-t-elle d'une grammaire universelle et rationnelle de notre époque informatisée? C'est seulement l'expérience pratique c'est-à-dire des essais très concrets, qui pourra répondre à ces questions.

A LA RECHERCHE D'UNE GRAMMAIRE UNIVERSELLE

NOTES

- ¹ Si on voulait développer syntaxe *et sémantique* (voir [1b]), on pourrait penser à la construction de modèles (au sens logique).
- ² Quoiqu'il s'agisse ici d'expressions bien formées, c'est-à-dire du pouvoir expressif, il ne faut pas non plus sous-estimer la capacité déductive (avec des prémisses additionnelles) des systèmes de premier ordre; capacité est bien comparable à celle des systèmes particuliers.
- ³ La Langue utilisée ici est micro-PROLOG (voir [1b]).

BIBLIOGRAPHIE

- [1] CHOMSKY, N., *Syntactic Structures*, La Haye, 1957.
- [1a] CLARK, K.L., et McCABE, F.G., *micro-PROLOG: Programming in Logic*, Englewood Cliffs, 1984.
- [1b] MONTAGUE, R., *Universal Grammar*, *Theoria*, Lund, 1970, vol. 36, pp. 373-398.
- [2] SAT-AMIKARO, *Nouveau cours rationnel et complet d'espéranto*, Paris, 1981.
- [3] STAHL, G., General Considerations about Modal Sentences, *Zeitsch. f. math. Logik*, Berlin, 1959, vol. V, pp. 280-290.
- [4] STAHL, G., Intensional Universes, *Philosophy and Phenom. Research*, Philadelphie, 1969, vol. XXX, n° 2, pp. 252-257.
- [5] STAHL, G., Termes temporels dans des systèmes fonctionnels, *Revue Philos. de la France et de l'étranger*, Paris, 1974, n° 3, pp. 293-303.
- [6] STAHL, G., Modal Models Corresponding to Models, *Zeitsch. f. math. Logik*, Berlin, 1974, vol. XX, pp. 407-410.
- [7] STAHL, G., Logical Treatment of the Relations of Knowing and Believing, *Philosophy and Phenom. Research*, Philadelphie, 1979, vol. XXXIX, pp. 511-523.
- [8] STAHL, G., La complétude d'un système temporel de *re*, *Logique et analyse*, Louvain, 1980, pp.365-373.

Gerold STAHL

- [9] STAHL, G., Quelques caractéristiques des modalités logiques, dans David, J. et Kleiber, G., *La notion sémantico-logique de modalité*, Metz, 1983, pp. 43-53.
- [10] STAHL, G., Quelques remarques sur la traduction assistée par ordinateur, *Semantics*, vol. 7, n° 2, Paris, 1983, pp. 67-79.

C.N.R.S (Paris)