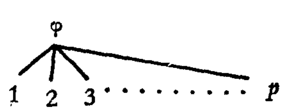
Jean Pierre Descles

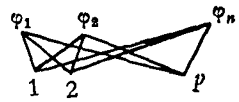
UN MODÈLE MATHÉMATIQUE D'ANALYSE TRANSFORMATIONNELLE SELON Z. S. HARRIS

Dans le projet général de Z. S. Harris (1968), nous avons à la base un ensemble K de phrases élémentaires (les phrases noyaux) et un ensemble Φ, défini pour chaque langue naturelle, d'opérateurs, de base. Une phrase est alors caractérisée par une unique décomposition (représentée par un treillis) obtenue par un produit des opérateurs de base et des phrases élémentaires de K. Ce qui caractérise donc une phrase, c'est un sous ensemble de K et un ensemble partiellement ordonné d'opérateurs qui agissent sur ce sous ensemble et ses transformés.

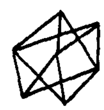
A partir de $T(\Phi)$, on peut construire la « théorie libre » de base Φ au sens de F. W. Lawvère (1963), retrouvé par J. Benabou (1968), repris par S. Eilenberg et J. B. Wright (1967). Une « théorie » est une catégorie mathématique particulière dont les objets sont les segments finis $[n]$ ($= \{1, 2, ..., n\}$ si $n \neq 0$ et Φ sinon) et les morphismes φ: $[n] \rightarrow [p]$ vérifient l'axiome de « théorie » (F. W. Lawvère, 1963). En employant la « théorie des graphes » (C. Berge, 1958) on peut donner une présentation plus intuitive de la construction de la « théorie libre ». A chaque opérateur de base φ de Φ, on associe un dendron (M. C. Barbault, J. P. Descles, 1972), c'est à dire un arbre étiqueté où tout noeud a l'ensemble de ses successeurs immédiats totalement ordonné. Un opérateur φ de poids p est représenté par un dendron de la forme:



On introduit deux opérations: 1) l'*intrication*: étant donnés n dendrons $\varphi_1, \varphi_2, ..., \varphi_n$ associés à n opérateurs de poids p (i.e. des opérateurs p-aires), on les intrique de manière à obtenir un opérateur complexe qui à p arguments associe un n-upple de valeurs. Chaque opérateur complexe est représenté par ce que nous appellons une treille (M. C. BARBAULT, J. P. DESCLES, 1970; 1972), de la forme:



et notée T^p_n; 2) la *greffe*: étant données deux treilles T^p_n et T^k_p, on greffe T^k_p aux terminaux de T^p_n pour obtenir la treille T^k_n:



$$\left.\begin{matrix} T^2_3 \\ T^3_2 \end{matrix}\right\} \; T^2_2 = T^2_3 \circ T^3_2$$

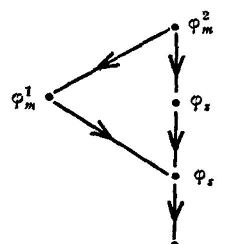
La « théorie libre » $T(\Phi)$ est un ensemble d'opérateurs qui contient: 1) l'ensemble des opérateurs de base Φ; 2) l'ensemble des treilles obtenues par intrication; 3) l'ensemble de toutes les applications ensemblistes des segments finis dont les projections et les permutations; 4) la composition par greffe de toutes les opérations précédentes. A chaque décomposition d'une phrase, on associe une treille qui représente un opérateur complexe construit à partir de l'ensemble partiellement ordonné des opérateurs de base qui entrent dans la décomposition d'une phrase. Cet opérateur complexe agit sur un sous ensemble ordonné de K pour donner une phrase complexe obtenue à partir des phrases élémentaires de K. $T(\Phi)$, construite à partir de Φ, sert de support à la formalisation des produits des opérateurs de base de Φ et à la décomposition transformationnelle de toute phrase. On obtient ainsi, par une construction algébrique, un ensemble d'opérateurs que l'on peut composer entre eux. Cette construction conduit à un traitement automatique des analyses transformationnelles. Et effet, à partir de l'approche définie par M. A. ARBIB et Y. GIVE'ON (1968) qui ont donné un modèle de calcul en parallèle directement associé à toute « théorie libre » et repris par nous dans M. C. BARBAULT, J. P. DESCLES (1970), on peut construire un automate qui produit (et analyse) les phrases d'une langue naturelle.

On définit alors la classe des grammaires transformationnelles GT par la donnée du triplet $< V, K, \Phi >$ où V est un ensemble de morphèmes, K est un sous ensemble de $V^\star$ (monoïde libre de base V), Φ un ensemble d'opérateurs de base d'une langue donnée. Une telle grammaire GT doit engendrer à partir de K l'ensemble des phrases de la langue et leurs analyses transformationnelles en termes d'opérateurs de base. D'autre part, la grammaire GT doit permettre de décomposer toute phrase en un ensemble partiellement ordonné d'opérateurs et de phrases élémentaires de K. Chaque décomposition transformationnelle est un opérateur complexe de l'ensemble $T(\Phi)$.

On peut ainsi donner un fragment d'une grammaire transformationnelle de l'anglais à partir de la décomposition proposée par Z. S. Harris (1968, p. 110). Harris donne pour l'anglais sept types d'opérateurs de base qui sont binaires ou unaires: les opérateurs qui sont des expansions de mots: φ_a (*strict; très strict*); les opérateurs φ_v portant sur le verbe (*il s'en va; il est parti (hier)*); les opérateurs portant sur les phrases φ_s (*je crois que...*); les connecteurs φ_c; les permutations φ_p (*je soupçonne cela; cela, je le soupçonne*); effacement et introduction de substitut φ_z; les opérateurs morphophonématiques φ_m.

Chaque opérateur complexe de $T(\Phi)$ qui est de poids p opère sur un p-uple de phrases élémentaires de K. Chacun des opérateurs T_n^p de $T(\Phi)$ fait passer d'un p-uple de phrases à un n-uple de phrases. Ainsi, $T(\Phi)$ qui opére sur K permet de: 1) construire l'ensemble des phrases S qui contient K; 2) structurer S par une série de relations (les transformations) à partir des opérateurs complexes de $T(\Phi)$. On est situé alors dans ce que Lawvère appelle le $T(\Phi)$-algèbre sur K.

Donnons un exemple du traitement de la phrase (Z. S. Harris, 1968, p. 133) *Frost reads smoothly* qui signifie « *Frost lit doucement* ». Dans $T(\Phi)$, on construit la treille associée à l'opérateur complexe:



treille

$$\begin{array}{l} [1] \\ \downarrow \varphi_m^2 \\ [2] \\ \downarrow < \varphi_m^1, \varphi_z > \\ [1] \\ \downarrow \varphi_s \\ [1] \end{array}$$

opérateur complexe

obtenue à partir des opérateurs de base: φ_s – *is smooth*; φ_z – effacement de *of anything*; φ_m^1 – *ly*; φ_m^2 – intonation de la phrase. Si l'on fait opérer cet opérateur complexe sur la phrase élémentaire *Frost reads anything* de K en suivant l'ordre partiel exprimé par la treille, on voit que les opérateurs φ_m^1 et φ_z opèrent en parallèle.

On obtient en définitive le sous-ensemble de S structuré par des transformations, que l'on peut représenter par le graphe orienté (ou treillis):

$$\begin{tikzcd}[row sep=small]
 & \textit{Frost reads smoothly} \\
\textit{Frost reads anything smoothly} \arrow[ur] & \textit{Frost's reading is smooth} \arrow[u] \\
 & \textit{Frost's reading of anything is smooth} \arrow[u] \arrow[ul] \\
 & \textit{Frost reads anything} \arrow[u]
\end{tikzcd}$$

Ce modèle mathématique pour être opératoire rend nécessaire d'expliciter, pour une langue donnée, les opérateurs de base de Φ. Z. S. Harris (1968) donne une méthode empirique pour dresser la liste de ces opérateurs. Les détails mathématiques et les preuves sont donnés dans M. C. Barbault, J. P. Descles, (1973), J. P. Descles (1973, à paraître) et M. A. Arbib, Y. Give'on (1968).

La formalisation d'un tel modèle permet de comparer ce modèle à d'autres modèles en particulier aux modèles de N. Chomsky et de S. K. Šaumjan (1972).

BIBLIOGRAPHIE

M. A. ARBIB, Y. GIVE'ON, *Algebra Automata*, I, II, dans « Information and Control », XII (1968).

M. C. BARBAULT, J. P. DESCLES, *Approches structurelles et catégorielles du système de transition*, Paris, 1970.

M. C. BARBAULT, J. P. DESCLES, *Transformations formelles et théories linguistiques*, Doc. Centre ling. quantitative n. 11, Paris, 1972.

M. C. BARBAULT, J. P. DESCLES, *Modèle mathématique transformationnel*, dans « T. A. information », (1973).

J. BENABOU, *Structures algébriques dans les catégories*, dans « Cahiers de topologie et géométrie différentielle », X (1968), 1.

C. BERGE, *Théorie des graphes et ses applications*, Paris, 1958.

J. P. DESCLES, *Théorie et T-algèbre-applications à la linguistique*, dans *Mathématiques et sciences humaines*, Paris, 1973 (à paraître).

S. EILENBERG, J. B. WRIGHT, *Automata in general algebras*, dans « Information and Control », XI (1967).

Z. S. HARRIS, *Mathematical Structures of Language*, New York, 1968.

F. W. LAWVÈRE, *Functorial semantic of algebraic theory*, dans « Proceedings of the National Academy of Sciences », 1963.

S. K. ŠAUMJAN, *Problèmes philosophiques de la théorie linguistique*, Moscou, 1972.