Catégories Morphologiques et analyse contextuelle dans
la linguistique algébrique

par Solomon Marcus

Une caractéristique du développement récent de la linguistique algébrique c'est l'existence de maintes théories parallèles, ayant pour objet une même réalité linguistique ou des faits assez semblables. Mais pour la plupart de ces théories on ne connaît pas les rapports exacts et on continue d'utiliser même des terminologies différentes pour des notions identiques, ce qui n'est pas de nature à favoriser le progrès de cette science. C'est justement la situation dans les études d'analyse algébrique contextuelle. Nous nous proposons de discuter, dans ce qui suit, les notions introduites par Dobrušin [1], [2] Sestier [11] et Sakai [10], notions qui s'avèrent intimement liées, parfois même équivalentes.

Considérons un vocabulaire fini $\underline{V}$. Toute suite finie d'éléments de $\underline{V}$ est une phrase (sur $\underline{V}$) ; Toute collection de phrases (sur $\underline{V}$) est un langage (sur $\underline{V}$). Il sera sous-entendu le fait que toutes les notions sont relatives à $\underline{V}$ et un langage $\underline{L}$ sur $\underline{V}$ et nous allons supprimer la référence explicite à ce fait.

Soient $\underline{a} \in \underline{V}$ et $\underline{b} \in \underline{V}$. Soit $\underline{L}$ un langage sur $\underline{V}$. On dit que $\underline{a}$ domine $\underline{b}$ et on écrit $\underline{a} \longrightarrow \underline{b}$ si pour tout couple de phrases $\underline{f}$ et $\underline{g}$ la relation $\underline{f}\underline{a}\underline{g} \in \underline{L}$ implique la relation $\underline{f}\underline{b}\underline{g} \in \underline{L}$. Si $\underline{L}$ est une langue naturelle, la relation $\underline{a} \longrightarrow \underline{b}$ a l'interprétation suivante. L'homonymie morphologique qui se manifeste parmi les formes flexionnelles du mot $\underline{a}$ est plus pauvre ou égale à celle qui se manifeste parmi les formes flexionnelles du mot $\underline{b}$. Par exemple, si $\underline{V}$ est le vocabulaire du français écrit et si $\underline{L}$ est la collection des phrases correctes du français écrit, alors, en prenant $\underline{a}$ = petit et $\underline{b}$ = mince, on constate que $\underline{a} \longrightarrow \underline{b}$, mais on n'a pas $\underline{b} \longrightarrow \underline{a}$ et ce fait est dû à ce que à des formes flexionnelles différentes de mince correspondent des formes flexionnelles différentes de petit, mais la réciproque n'est pas vraie, car aux formes petit et petite correspond la même forme mince. Si l'on a, à la fois, $\underline{a} \longrightarrow \underline{b}$ et $\underline{b} \longrightarrow \underline{a}$, on écrit $\underline{a} \longleftrightarrow \underline{b}$ et il est facile de voir que $\longleftrightarrow$ est une relation d'équivalence dans $\underline{V}$. La classe d'équivalence de $\underline{a}$ sera notée par $\underline{S}(\underline{a})$; c'est la classe de distribution du mot $\underline{a}$.

Par exemple, dans l'exemple du français écrit envisagé ci-dessus, on a $\text{petit} \longleftrightarrow \text{grand}$, donc $\text{grand} \in S(\text{petit})$. Un mot $\underline{a}$ est dit initial si pour $\underline{b} \longrightarrow \underline{a}$ on a toujours $\underline{b} \in S(\underline{a})$. R.L. Dobrušin a introduit la notion de catégorie morphologique (Dobrušin dit : grammaticale) élémentaire engendrée par un mot initial $\underline{a}$; c'est l'ensemble $\mathcal{C}_j(\underline{a}) = \{ \underline{b} ; \underline{a} \longrightarrow \underline{b} \}$. L'interprétation de cette notion est la suivante : $\mathcal{C}_j(\underline{a})$ est l'ensemble des mots qui possèdent toutes les valeurs morphologiques de $\underline{a}$. Par exemple, dans le français écrit $\underline{a} = \text{petit}$ est un mot initial, tandis que $\mathcal{C}_j(\underline{a}) = \{ \text{petit}, \text{grand}; \text{mince}, \text{gros}, \dots \}$ est l'ensemble des adjectifs qualificatifs qui sont au masculin singulier. Naturellement, il peut arriver que certains mots de $\mathcal{C}_j(\underline{a})$ possèdent, en outre, des valeurs étrangères à $\underline{a}$, comme on constate aussi sur l'exemple ci-dessus : le mot mince possède aussi la valeur du féminin, tandis que le mot gros possède aussi la valeur du pluriel.

Mais la notion de catégorie morphologique élémentaire ne couvre pas toutes les catégories morphologiques de la linguistique. Par exemple, la catégorie du singulier, celle du féminin, celle de l'adjectif qualificatif et beaucoup d'autres ne correspondent à aucune catégorie morphologique élémentaire. Ce fait nous a déterminé d'introduire une certaine extension de la notion de Dobrušin (voir [3], [4], [5] et le chapitre V de [8]). Si $A \subseteq V \supseteq B$ et si $\underline{a} \longrightarrow \underline{b}$ pour chaque $\underline{a} \in A$ et pour chaque $\underline{b} \in B$, on dit que A domine B et on écrit $A \longrightarrow B$. Posons $A_1 = \{ \underline{a}; A \longrightarrow \underline{a} \}$. L'ensemble $\mathcal{C}_j(A) = A \cup A_1$ est, par définition, la catégorie morphologique engendrée par A . Dans le cas particulier où $A = \{ \underline{a} \}$ et $\underline{a}$ est un mot initial, on obtient la notion de catégorie morphologique élémentaire, due à Dobrušin. L'intérêt linguistique de la notion que nous venons de définir vient du fait qu'elle couvre toutes les catégories morphologiques intuitives de la linguistique. Ce fait est la conséquence des deux propositions suivantes :

1° - La réunion de deux catégories morphologiques est aussi une catégorie morphologique (voir la Proposition de [4], P. 328, où les catégories morphologiques sont appelées catégories grammaticales).

2° - Toute catégorie morphologique intuitive correspond à une réunion de catégories morphologiques élémentaires (ce qui justifie la dénomination

de catégorie morphologique élémentaire ; c'est l'atome morphologique, la cellule qui se trouve à la base de toute catégorie morphologique). Prenons, par exemple, la catégorie morphologique du singulier des adjectifs qualificatifs du français écrit. Les mots appartenant à cette catégorie forment un ensemble qui n'est autre chose que la réunion des catégories morphologiques élémentaires $\mathcal{C}_f(\text{petit})$ et $\mathcal{C}_f(\text{petite})$. En ce qui concerne la catégorie morphologique des adjectifs qualificatifs, elle est une réunion de quatre catégories morphologiques élémentaires, à savoir $\mathcal{C}_f(\text{petit})$, $\mathcal{C}_f(\text{petite})$, $\mathcal{C}_f(\text{petits})$, $\mathcal{C}_f(\text{petites})$. En ce qui concerne la réunion de deux catégories morphologiques élémentaires distinctes, elle n'est jamais une catégories morphologique élémentaire. En effet, soient $\mathcal{C}_f(a)$ et $\mathcal{C}_f(b)$ deux catégories morphologiques élémentaires distinctes, et supposons qu'il existe un mot initial $\underline{c}$ tel que $\mathcal{C}_f(c) = \mathcal{C}_f(a) \cup \mathcal{C}_f(b)$ (les mots $\underline{a}$ et $\underline{b}$ sont aussi initiaux, par la définition même de la notion de catégorie morphologique élémentaire). On a donc $\underline{c} \rightarrow \underline{a}$ et $\underline{c} \rightarrow \underline{b}$. Cela implique, en vertu du fait que les mots $\underline{a}$ et $\underline{b}$ sont initiaux, $\underline{a} \in S(c)$ et $\underline{b} \in S(c)$, donc $\underline{a} \leftrightarrow \underline{b}$. Mais on obtient ainsi une contradiction avec l'hypothèse que $\mathcal{C}_f(a) \neq \mathcal{C}_f(b)$; donc la réunion de $\mathcal{C}_f(a)$ et $\mathcal{C}_f(b)$ n'est plus une catégorie morphologique élémentaire. Mais cette réunion est toujours (en vertu du 1°) une catégorie morphologique, ce qui prouve l'avantage de cette dernière notion.

Passons maintenant au point de vue de Sestier [11] qui sera présenté ici avec certaines modifications de terminologie et de notation. Tout couple ordonné $\langle f, g \rangle$ de phrases (sur V) est un contexte (sur V). On dit que le contexte $\langle f, g \rangle$ accepte la phrase $\underline{h}$ (par rapport à un langage $\underline{L}$ sur $\underline{V}$) si $fgh \in \underline{L}$. Désignons par $\alpha(h)$ l'ensemble des contextes qui acceptent la phrase h et par $\alpha^{-1}(\gamma)$ l'ensemble des mots acceptés par le contexte γ . Soit maintenant un ensemble X V , et un ensemble Y de contextes. Posons

$$\alpha(X) = \overline{\alpha(X)} \quad , \quad \alpha^{-1}(Y) = \overline{\alpha^{-1}(Y)}$$

$$\Psi(X) = \alpha^{-1}(\alpha(X)) \quad \text{et} \quad \Psi(Y) = \alpha(\alpha^{-1}(Y))$$

L'ensemble $\Psi(X)$ est la fermeture de Sestier de l'ensemble X , tandis que $\Psi(Y)$ est la fermeture de Sestier de l'ensemble Y .

Dans le premier cas, la fermeture est un ensemble de mots, tandis que dans le deuxième elle est un ensemble de contextes.

Comme nous l'avons montré ailleurs (propositions 1, 2 et 3 de [7] , on a, pour tout ensemble X de mots, l'inclusion $\mathcal{C}_f(X) \subseteq \mathcal{P}(X)$; il y a des cas où l'inclusion est stricte, mais si X se réduit à un seul mot, on en a toujours l'égalité. En particulier, il s'ensuit que la catégorie morphologique élémentaire engendrée par un mot initial coïncide avec la fermeture de Sestier de ce mot.

Il est intéressant de faire une confrontation entre les catégories morphologiques et les fermetures de Sestier car les unes et les autres ont des significations linguistiques très importantes. Voici d'abord quelques analogies (X est une partie quelconque de V) :

a) $X \subseteq \mathcal{C}_f(X)$;	a') $X \subseteq \mathcal{P}(X)$;
b) $\mathcal{C}_f(\mathcal{C}_f(X)) = \mathcal{C}_f(X)$;	b') $\mathcal{P}(\mathcal{P}(X)) = \mathcal{P}(X)$;
c) $V - \mathcal{C}_f(X) \subseteq \mathcal{C}_f(V - X)$;	c') $V - \mathcal{P}(X) \subseteq \mathcal{P}(V - X)$;
d) il y a des cas où l'inclusion a) est stricte et il y en a d'autres où l'on a l'égalité ;	d') il y a des cas où l'inclusion a') est stricte et il y en a d'autres où l'on a l'égalité ;
e) il y a des cas où l'inclusion c) est stricte et il y en a d'autres où l'on a l'égalité ;	e') il y a des cas où l'inclusion c') est stricte et il y en a d'autres où l'on a l'égalité ;
f) afin que X soit une catégorie morphologique il faut et il suffit que $\mathcal{C}_f(X) = X$	f') afin que X soit une fermeture de Sestier il faut et il suffit que $\mathcal{P}(X) = X$

Les propositions a) et a') sont évidentes. La proposition b) est le théorème 5 de [4] , tandis que la proposition b') a été donnée dans [12] (voir aussi la proposition 18 de [12]). Les propositions c) et c') sont, respectivement, les propositions 23 et 19 de [12] . Les propositions d) et d') sont évidentes. Les propositions e) et e') résultent respectivement des propositions 24 et 20 de [12] . Les propositions f) et f') sont respectivement des conséquences de b) et b').

Il y a probablement maintes autres analogies entre $\mathcal{U}$ et φ et on pourrait chercher ainsi l'analogie pour φ des théorèmes de [3], [4], [5] et du chapitre V de [8]. Mais nous voulons attirer l'attention sur quelques propriétés qui montrent les différences de nature algébrique et de signification linguistique des deux opérateurs $\mathcal{U}$ et φ (A et B sont deux sous-ensembles de V) :

g) il y a des cas où $A \subset B$;
mais aucun des ensembles $\mathcal{U}(A)$
et $\mathcal{U}(B)$ n'est contenu dans
l'autre ;

h) la réunion de deux catégories
morphologiques est aussi une
catégorie morphologique ;

i) l'intersection de deux caté-
gories morphologiques n'est pas
toujours une catégorie morpholo-
gique ;

j) il y a des cas où $\mathcal{U}(A \cap B)$
n'est pas contenu dans
 $\mathcal{U}(A) \cap \mathcal{U}(B)$;

k) il y a des ensembles A et B
tels que $\mathcal{U}(A \cup B) \subset \mathcal{U}(A) \cup \mathcal{U}(B)$

l) Si $A - B \neq 0 \neq B - A$,
alors $\mathcal{U}(A) = \mathcal{U}(B)$ si et
seulement s'il existe un mot a
tel que $A \cup B \rightarrow S(a) \supseteq A \Delta B$,
où Δ signifie la différence
symétrique.

$$g') A \subseteq B \Rightarrow \varphi(A) \subseteq \varphi(B) ;$$

h') la réunion de deux fermetures de Sestier
n'est pas toujours une fermeture de Sestier

i') l'intersection de deux fermetures de
Sestier est aussi une fermeture de
Sestier ;

$$j') \varphi(A \cap B) \subseteq \varphi(A) \cap \varphi(B) ;$$

$$k') \varphi(A) \cup \varphi(B) \subseteq \varphi(A \cup B)$$

$$l') \varphi(A) = \varphi(B) \text{ si et seulement si } \alpha(A) = \alpha(B).$$

La proposition g) résulte de l'exemple suivant. Soit
 $V = \{a, b, c, d, e\}$, $L = \{abc, adc, ebc, edc\}$, $A = \{a\}$, $B = \{a, b, c\}$.
On a $A \subset B$, $\mathcal{U}(A) = \{a, e\}$, $\mathcal{U}(B) = \{a, b, c\}$, donc aucun des ensembles
 $\mathcal{U}(A)$ et $\mathcal{U}(B)$ n'est contenu dans l'autre. La proposition g') est la proposi-
tion 12 de [12]. La proposition h) a été établie dans [4], p. 327-328. Quant
à la proposition h'), il faut remarquer qu'elle est équivalente (en vertu de la

proposition b')) à l'existence de deux ensembles A et B qui ne satisfont pas l'égalité $\varphi(\varphi(A) \cup \varphi(B)) = \varphi(A) \cup \varphi(B)$. C'est justement le cas, comme le montre l'exemple suivant, dû à Grzegorz Rozenberg (communication orale). Soit $V = \{a, b, c, d, e, f, g, \}$, $L = \{cac, cdc, gbg, geg, cfc, gfg\}$. Posons $A = \{a, d\}$, $B = \{b, e\}$. On a $\varphi(A) = \{a, d, f\}$, $\varphi(B) = \{b, e, f\}$
 $\varphi(A) \cup \varphi(B) = \{a, b, d, e, f\}$,
 $\varphi(\varphi(A) \cup \varphi(B)) = \alpha^{-1}(\alpha(\{a, b, d, e, f, \})) = \alpha^{-1}(0) = v$

Ce qui rend plus intéressant cet exemple c'est le fait que A et B sont ici des classes de distribution ; en effet, on a $A = S(a)$ et $B = S(b)$. Il s'ensuit donc que, non seulement la réunion de deux fermetures de Sestier n'est pas toujours une fermeture de Sestier, mais elle ne l'est ni même dans le cas particulier où les deux fermetures envisagées sont des catégories morphologiques élémentaires (on sait, d'après la proposition 3 de [7], que $\mathcal{C}(S(x)) = \varphi(S(x))$, quel que soit le mot x). Mais nous avons vu que les réunions finies de catégories morphologiques élémentaires donnent justement les catégories morphologiques intuitives des langues naturelles ; il s'ensuit donc que ces catégories intuitives ne correspondent pas toujours à des fermetures de Sestier, ce qui met la notion de fermeture due à Sestier dans une situation d'infériorité par rapport à la notion de catégorie morphologique (si l'on envisage le point de vue morphologique). D'autre part, du point de vue de l'analyse contextuelle, c'est la fermeture de Sestier qui semble plus naturelle et plus utile.

La proposition i) résulte du même exemple utilisé pour prouver la proposition g). En effet, on a, dans cet exemple, $\mathcal{C}(A) \cap \mathcal{C}(B) = \{a\}$, mais il existe aucun ensemble C tel que $\mathcal{C}(C) = \{a\}$. En effet, on a $\mathcal{C}(\{a\}) = \{a, e\}$, donc $\mathcal{C}(\{a\}) \neq \{a\}$ et, en vertu de la proposition f), $\{a\}$ n'est pas une catégorie morphologique. En ce qui concerne la proposition i'), il faut remarquer qu'elle est équivalente à l'assertion $\varphi(\varphi(A) \cap \varphi(B)) = \varphi(A) \cap \varphi(B)$. Voici une démonstration rapide de cette égalité, due à R. S. Waligorski (communication orale). En utilisant les propositions j') et b'), on a $\varphi(\varphi(A) \cap \varphi(B)) \subseteq \varphi(\varphi(A)) \cap \varphi(\varphi(B)) = \varphi(A) \cap \varphi(B)$. D'autre part, en vertu de la proposition a'), on a $\varphi(A) \cap \varphi(B) \subseteq \varphi(\varphi(A) \cap \varphi(B))$. La proposition j) a été établie par E. Toma Marcu (voir la proposition 22 du [12]). Le même auteur a démontré la proposition j') (voir la proposition 13 de [12]). La proposition k) résulte de l'exemple suivant. Soient $V = \{a, b, c, d, e, f, g, h, \}$

$L = \{abc, adc, efg, ehg\}$. Posons $A = \{b\}$, $B = \{f\}$. On a $\mathcal{U}_f(A) = \{b, d\}$,
 $\mathcal{U}_f(B) = \{f, h\}$, $\mathcal{U}_f(A \cup B) = \mathcal{U}_f(\{b, f\}) = \{b, f\}$, donc $\mathcal{U}_f(A \cup B) \subset \mathcal{U}_f(A) \cup \mathcal{U}_f(B)$
 La proposition k') est justement la proposition 15 de [12]. La proposition l) a
 été démontrée dans [3] (voir aussi le chapitre V de [8]) tandis que la proposi-
 tion l') est la proposition 18 de [12].

On constate donc qu'il y a une certaine dualité entre les opérateurs
 G et $\mathcal{U}$, en ce qui concerne les opérateurs de réunion et d'insertion (voir les
 propositions h) et h'), i) et i')) mais en général l'opérateur $\mathcal{U}$ présente une
 certaine supériorité de nature algébrique (voir, par exemple, la propriété de
 monotonie de l'opérateur $\mathcal{U}$). Il serait intéressant d'étudier le comportement
 réciproque de ces deux opérateurs, par exemple leur superposition, la distributivité
 de l'un par rapport à l'autre, etc...

Dans un travail récent, Itiroo Sakai développe une analyse contex-
 tuelle qui en beaucoup de points s'approche des notions ci-dessus, dès que ces
 notions sont envisagées non seulement pour les mots (c'est-à-dire pour les
 éléments de V), mais aussi pour les phrases (c'est-à-dire pour les éléments du
 langage universel sur V), il introduit la notion de voisinage complet d'une
 phrase f , qui revient à l'ensemble $\mathcal{A}(f)$, et la notion de voisinage élémentaire,
 qui revient à un voisinage complet qui est une classe de distribution, c'est-à-
 dire qui est de la forme $S(f)$. Dans [8] nous avons établi certaines propositions qui
 lient les notions de Sakai et celles ci-dessus, mais il reste encore à approfondir
 cette liaison. Des aspects nouveaux de certaines des notions ci-dessus sont
 étudiés par Revzin [9] et par Marcus [6].

B i b l i o g r a f i e

1. R.L. Dobruşin Elementarnaja grammatičeskaja kategorija
Bjulleten Obdeinenija po probleman maşinnago
perevoda, 1957, nr. 5, p. 19-21
2. R.L. Dobruşin Matematičeskie metody v lingvistike. Priloženie
Matematičeskoe prosvescenie, vol. 6, 1961 p. 52-59
3. S. Marcus sur un modèle logique de la catégorie grammaticale
élémentaire, I. Revue de mathématiques pures et appli-
quées, vol. 7, 1962; nr. 1, p. 91-107
4. S. Marcus Sur un modèle logique de la catégorie grammaticale
élémentaire, II. Zeitschrift für Mathematische Logik
und Gröndlagen der Mathematik, vol. 8, 1962, nr. 3-4,
p. 323-329.
5. S. Marcus Ob odnoi logičeskoj modeli elementarnoi grammatičeskoj
kategorii, III, Revue de Mathématiques pures et
appliquées, vol. 7, 1962, nr. 4., p. 683-691.
6. S. Marcus Analyse contextuelle. Zeitschrift für Phonetik,
Sprachwissenschaft und Kommunikationsforschung,
vol. 18, 1965, nr. 3, p. 301-313
7. S. Marcus Catégories de Dobruşin, fermetures de Sestier et
voisinages de Sakai, Glosse, vol. 1, 1967, nr.1.
8. S. Marcus Introduction mathématique à la linguistique structu-
rale. Dunod, Paris; 1967
9. I.I. REVZIN Metod modelirovanija i typologia slavjanskich
jazykov, Izd. Nauka, Moskva, 1967
10. I. Sakai Some mathematical aspects of syntactic description
(Preprint). International Conference of Computational
linguistics, New-York, 1965, nr.21
11. A.Sestier Contribution à une théorie ensembliste des classifications
linguistiques, actes du Premier Congrès de l'AFCAL, 1960
Paris, 1961, P. 293-305
12. E. Toma Marcu Proprietăţi ale închiderilor Sestier din teoria
algebrică a gramaticii. Studii şi cercetări matematice
vol.12, 1967.