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

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

Par exemple, dans l'exemple du français écrit envisagé ci-dessus, on a petit $\longleftrightarrow$ grand, donc grand $\in S(\text{petit})$. Un mot *a* est dit initial si pour $b \longrightarrow a$ on a toujours $b \in S(a)$. R.L. Dobrušin a introduit la notion de catégorie morphologique (Dobrušin dit : grammaticale) élémentaire engendrée par un mot initial *a* ; c'est l'ensemble $\mathcal{G}(a) = \{b\ ;\ a \longrightarrow b\}$. L'interprétation de cette notion est la suivante : $\mathcal{G}(a)$ est l'ensemble des mots qui possèdent toutes les valeurs morphologiques de *a*. Par exemple, dans le français écrit a = petit est un mot initial, tandis que $\mathcal{G}(a) = \{$ petit, grand; mince, gros, ... $\}$ est l'ensemble des adjectifs qualificatifs qui sont au masculin singulier. Naturellement, il peut arriver que certains mots de $\mathcal{G}(a)$ possèdent, en outre, des valeurs étrangères à *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 $a \longrightarrow b$ pour chaque $a \in A$ et pour chaque $b \in B$, on dit que A domine B et on écrit $A \longrightarrow B$. Posons $A_1 = \{a;\ A \longrightarrow a\}$. L'ensemble $\mathcal{G}(A) = A \cup A_1$ est, par définition, la catégorie morphologique engendrée par A. Dans le cas particulier où $A = \{a\}$ et *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 $\mathcal{C}$(<u>petit</u>) et $\mathcal{C}$ (petite). En ce qui concerne la catégorie morphologique des adjectifs qualificatifs, elle est une réunion de quatre catégories morphologiques élémentaires, à savoir $\mathcal{C}$(<u>petit</u>) , $\mathcal{C}$ (<u>petite</u>) , $\mathcal{C}$ (<u>petits</u>) , $\mathcal{C}$ (<u>petites</u>) . En ce qui concerne la réunion de deux catégories morphologiques élémentaires distinctes, elle n'est jamais une catégories morphologique élémentaire. En effet, soient $\mathcal{C}$(a) et $\mathcal{C}$(b) deux catégories morphologiques élémentaires distinctes, et supposons qu'il existe un mot initial <u>c</u> tel que $\mathcal{C}(c) = \mathcal{C}(a) \cup \mathcal{C}(b)$ (les mots <u>a</u> et <u>b</u> sont aussi initiaux, par la définition même de la notion de catégorie morphologique élémentaire). On a donc <u>c</u> $\longrightarrow$ <u>a</u> et <u>c</u> $\longrightarrow$ <u>b</u> . Cela implique, en vertu du fait que les mots <u>a</u> et <u>b</u> sont initiaux, <u>a</u> $\in$ S(c) et <u>b</u> $\in$ S (c), donc <u>a</u> $\longleftrightarrow$ <u>b</u> . Mais on obtient ainsi une contradiction avec l'hypothèse que $\mathcal{C}(a) \neq \mathcal{C}(b)$; donc la réunion de $\mathcal{C}$(a) et $\mathcal{C}$(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 <u>h</u> (par rapport à un langage <u>L</u> sur <u>V</u>) si $fgh \in L$. 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

$$\alpha(X) = \bigcap_{x \in X} \alpha(x) \quad , \quad \alpha^{-1}(Y) = \bigcap_{y \in Y} \alpha^{-1}(y)$$

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

L'ensemble φ (X) est la fermeture de Sestier de l'ensemble X , tandis que ψ (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 $\mathcal{G}(X) \subseteq \varphi(X)$; 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) $X \subseteq \mathcal{G}(X)$;
b) $\mathcal{G}(\mathcal{G}(X)) = \mathcal{G}(X)$;
c) $V - \mathcal{G}(X) \subseteq \mathcal{G}(V - X)$;
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 ;
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 $\mathcal{G}(X) = X$

a') $X \subseteq \varphi(X)$;
b') $\varphi(\varphi(X)) = \varphi(X)$;
c') $V - \varphi(X) \subseteq \varphi(V - X)$;
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 ;
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 fermeture de Sestier il faut et il suffit que $\varphi(X) = X$

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 $\mathcal{G}$ et φ et on pourrait chercher ainsi l'analogue 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 $\mathcal{G}$ 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 $\mathcal{G}(A)$ et $\mathcal{G}(B)$ n'est contenu dans l'autre ;

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

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

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

i) l'intersection de deux catégories morphologiques n'est pas toujours une catégorie morphologique ;

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

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

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

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

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

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

l') $\varphi(A) = \varphi(B)$ 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 $A \subset B$, $\mathcal{G}(A) = \{a, e\}$, $\mathcal{G}(B) = \{a, b, c\}$, donc aucun des ensembles $\mathcal{G}(A)$ et $\mathcal{G}(B)$ n'est contenu dans l'autre. La proposition g') est la proposition 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 $\mathcal{G}(S(x)) = \varphi(S(x))$, 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, $\mathcal{G}(A) \cap \mathcal{G}(B) = \{a\}$, mais il existe aucun ensemble C tel que $\mathcal{G}(C) = \{a\}$. En effet, on a $\mathcal{G}(\{a\}) = \{a, e\}$, donc $\mathcal{G}(\{a\} \neq \{a\}$ 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 $\mathcal{G}(A) = \{b, d\}$, $\mathcal{G}(B) = \{f, h\}$, $\mathcal{G}(A \cup B) = \mathcal{G}(\{b, f\}) = \{b, f\}$, donc $\mathcal{G}(A \cup B) \subset \mathcal{G}(A) \cup \mathcal{G}(B)$
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 proposition l') est la proposition 18 de [12] .

On constate donc qu'il y a une certaine dualité entre les opérateurs G et φ , 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 φ présente une certaine supériorité de nature algébrique (voir, par exemple, la propriété de monotonie de l'opérateur φ). 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 contextuelle 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 $\alpha(f)$, 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 p h i e

1. R.L. Dobrušin — Elementarnaja grammatičeskaja kategorija
Bjulleten Obdeinenija po probleman mašinnago perevoda, 1957, nr. 5, p. 19-21

2. R.L. Dobrušin — 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 appliquées, 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 Grundlagen der Mathematik, vol. 8, 1962, nr. 3-4, p. 323-329.

5. S. Marcus — Ob odnoi logičeskoi modeli elementarnoi grammatičeskoi 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. 3o1-313

7. S. Marcus — Catégories de Dobrušin, fermetures de Sestier et voisinages de Sakai, Glosse, vol. 1, 1967, nr.1.

8. S. Marcus — Introduction mathématique à la linguistique structurale. Dunod, Paris; 1967

9. I.I. REVZIN — Metod modelirovanija i typologia slavianskich 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 Congrés de l'AFCAL, 1960 Paris, 1961, P. 293-305

12. E. Toma Marcu — Propiretăţi ale închiderilor Sestier din teoria algebrică a gramaticii. Studii şi cercetări matematice vol.12, 1967.