

# Le groupe de Hotz

Oleksandr Aksenov

11 février 2007

Le définition classique du langage engendré par la grammaire utilise les règles dans le sens de dérivation : de gauche à droite. Or, on peut aussi former les mots en mélangeant la dérivation et la réduction (le procédé inverse). En général, ce procédé de dérivation-réduction élargit les possibilités, et les propriétés de la congruence qu'on obtient dans le monoïde libre dépendent de la grammaire même pour des grammaires équivalentes. Un moyen d'obtenir un invariant qui ne dépend que du langage engendré est de passer aux quotients dans le groupe libre. Ce sont les groupes de Hotz [1]. Pour certaines grammaires, l'application des règles dans les deux sens ne modifie pas le langage. Ces grammaires sont appelées *grammaires NTS*.

**Définition 1.** Soit  $G = (A, V, P)$  une grammaire. Si  $u, v$  sont deux mots sur  $(A+V)$ , on note  $u \leftrightarrow v$  si on a  $u \rightarrow v$  ou  $v \rightarrow u$ . La clôture réflexive et transitive de la relation  $\leftrightarrow$  sera notée  $\overset{*}{\leftrightarrow}$  ou  $[P]$ . Il s'agit en effet de la plus fine congruence des monoïdes qui contient les règles de  $G$ . On appelle alors le *monoïde de Hotz* associé le monoïde quotient  $V^*/[P]$  du monoïde libre  $V^*$  par la congruence  $\overset{*}{\leftrightarrow}$ .

On s'intéressera au lien entre cette congruence et le langage engendré par la grammaire (quand les règles sont appliquées comme d'habitude, de gauche à droite). Dans ce cadre, les grammaires considérées seront le plus souvent réduites.

Voici un exemple.

*Exemple 2.* Considérons le langage  $L = a + a^+b$  avec  $A = \{a, b\}$  et deux grammaires différentes qui l'engendrent.

1. La grammaire définie par  $P = \{S \rightarrow a + Tb; T \rightarrow a + Ta\}$ . La congruence  $[P]$  est alors engendrée par  $S \leftrightarrow a \leftrightarrow T \overset{*}{\leftrightarrow} ab \overset{*}{\leftrightarrow} a^2$ .

2. La grammaire définie par  $P = \{S \rightarrow a + Tab; T \rightarrow 1 + Ta\}$ . On obtient alors

$$\begin{aligned} S &\leftrightarrow a, \\ Ta &\leftrightarrow T \leftrightarrow 1, \text{ d'où} \\ a &\overset{*}{\leftrightarrow} 1, \\ Tab &\leftrightarrow S \overset{*}{\leftrightarrow} 1, \\ b &\overset{*}{\leftrightarrow} 1. \end{aligned}$$

Le monoïde de Hotz est ici trivial.

On voit que les monoïdes obtenus ne sont pas isomorphes malgré le fait que ces grammaires engendrent le même langage. Ce problème disparaît par contre si on travaille avec des groupes.

**Définition 3.** Soit  $G$  une grammaire, et  $u, v \in F(V \cup \bar{V} \cup A \cup \bar{A})$ . On note alors  $u \leftrightarrow v$  s'il existe  $u_1, v_1, w, w' \in (V \cup \bar{V} \cup A \cup \bar{A})^*$  tels que  $u = wu_1w'$  et  $v = wv_1w'$  dans le groupe libre, et on est dans un des 4 cas suivants :

- $u_1 \in V$  et  $u_1 \rightarrow v_1$  est une règle de  $G$  ;
- $v_1 \in V$  et  $v_1 \rightarrow u_1$  est une règle de  $G$  ;
- $u_1 \in \bar{V}$  et  $\bar{u}_1 \rightarrow v_1^{-1}$  est une règle de  $G$  ;
- $v_1 \in \bar{V}$  et  $\bar{v}_1 \rightarrow u_1^{-1}$  est une règle de  $G$ .

On note  $\overset{*}{\leftrightarrow}$  la clôture réflexive et transitive de la relation  $\leftrightarrow$ . C'est aussi la plus fine congruence de groupe qui contient les règles de  $G$ , elle sera donc également notée  $[P]$ . On appelle *groupe de Hotz* associé à  $G$  le groupe  $\mathcal{H}(G) = F(V \cup \bar{V} \cup A \cup \bar{A})/[P]$ .

Dans l'exemple précédent le groupe de Hotz était trivial pour les deux grammaires. On va montrer que ce résultat (le groupe de Hotz ne dépend que du langage engendré par la grammaire à isomorphisme près) est assez général.

**Théorème 4.** Soit  $G$  une grammaire réduite et  $L$  le langage engendré par  $G$ . Le groupe de Hotz associé à  $G$  est alors isomorphe au groupe  $\mathcal{C}(L) = F(A \cup \bar{A})/[L \times L]$ , où  $[L \times L]$  est la plus fine congruence de groupe qui met  $L$  dans une seule classe de congruence.

*Démonstration.* Notons  $\equiv$  la congruence modulo  $[L \times L]$  définie sur  $F(A \cup \bar{A})$  et montrons que pour tous  $u, v \in F(A \cup \bar{A})$  on a

$$u \equiv v \Leftrightarrow u \overset{*}{\leftrightarrow} v.$$

En effet, soient  $u, v \in L$ . On a alors  $S \overset{*}{\rightarrow} u$  et  $S \overset{*}{\rightarrow} v$ , donc  $u \overset{*}{\leftrightarrow} v$ . Donc, la congruence modulo  $[L \times L]$  est plus fine que celle modulo  $[P]$ .

Réciproquement, soient deux mots  $u, v \in L_G(X)$  où  $X$  est une variable de la grammaire  $G$ . Comme  $G$  est réduite, il existe  $w, w' \in A^*$  tels que  $wXw' \in \widehat{L}_G(S)$ , d'où  $www' \equiv wv_1w'$ , et, comme  $\equiv$  est une congruence de groupe, on obtient  $u \equiv v$ .

Considérons un morphisme de groupes  $\phi$  de  $F(V \cup \bar{V} \cup A \cup \bar{A})$  dans  $F(A \cup \bar{A})$  tel que pour chaque terminal  $a$  on ait  $\phi(a) = a$ , et pour chaque variable  $X$  on ait  $\phi(X) \in L_G(X)$ . Pour chaque règle  $X \rightarrow \alpha$  de  $G$  on a alors (d'après ce qui précède)  $\phi(X) \equiv \phi(\alpha)$ . Par conséquent,  $u \leftrightarrow v$  implique  $\phi(u) \equiv \phi(v)$  quelque soient  $u, v \in F(V \cup \bar{V} \cup A \cup \bar{A})$ . Donc  $u \overset{*}{\leftrightarrow} v$  implique  $\phi(u) \equiv \phi(v)$ . Si, de plus,  $u, v \in F(A \cup \bar{A})$ , on obtient  $u \equiv v$ . Cela montre l'équivalence énoncée.

D'après cette équivalence, on peut définir l'application  $\check{\phi} : F(V \cup \bar{V} \cup A \cup \bar{A})/[P] \rightarrow F(A \cup \bar{A})/[L \times L]$  qui à chaque classe d'équivalence  $L'$  de  $F(V \cup \bar{V} \cup A \cup \bar{A})$  modulo  $\overset{*}{\leftrightarrow}$  associe la classe d'équivalence de  $F(A \cup \bar{A})$  incluse dans  $L'$ . On voit que  $\check{\phi}$  est un isomorphisme de groupes.  $\square$

*Exemple 5.* Pour la grammaire définie par  $P = \{S \rightarrow 1 + aSb\}$ , la congruence  $\overset{*}{\leftrightarrow}$  est engendrée par  $S \overset{*}{\leftrightarrow} 1 \overset{*}{\leftrightarrow} ab$ . Le groupe de Hotz associé est donc isomorphe à  $\mathbb{Z}$  par le morphisme  $w \rightarrow [w]_a - [w]_b$ .

Un moyen d'obtenir une correspondance entre le langage engendré par la grammaire et le monoïde de Hotz consiste à se restreindre à des grammaires particulières.

**Définition 6.** Une grammaire  $G$  est dite à *non terminaux séparés* (grammaire *NTS* en abrégé) si pour chaque variable  $X$  de  $G$  et chaque  $x \in (V \cup A)^*$  on a

$$X \overset{*}{\rightarrow} x \Leftrightarrow X \overset{*}{\leftarrow} x.$$

Un langage *NTS* est un langage généré par une grammaire *NTS* à partir d'un ensemble de ses variables pris comme axiomes.

Voici quelques exemples simples de grammaires *NTS*.

*Exemple 7. 1.* La grammaire définie par  $A = \{a, b\}, V = \{S\}, P = \{S \rightarrow SS; S \rightarrow a; S \rightarrow b\}$  est *NTS*. En effet,  $\widehat{L}_G(S) = \{a, b, S\}^+$  et on n'a pas  $\epsilon \overset{*}{\leftrightarrow} S$ .

*2.* La grammaire définie par  $P = \{S \rightarrow 1 + aSb\}$  n'est pas *NTS*. Par exemple,  $abab \notin \widehat{L}_G(S)$  mais on a la dérivation-réduction

$$S \rightarrow ab = ab1 \leftarrow abS \rightarrow abab.$$

*3.* Par contre, le langage  $\{a^n b^n | n \in \mathbb{N}^*\}$  reconnu par cette grammaire est un langage *NTS*. Un exemple de grammaire *NTS* qui l'engendre est la grammaire avec les règles

$$S \rightarrow aSb ; S \rightarrow ab.$$

On a ici  $\widehat{L}_G(S) = \{a^n S b^n | n \in \mathbb{N}\}$ .

*4.* Les grammaires de l'exemple 2 ne sont pas *NTS*. Par contre, le langage qu'elles engendrent l'est. Il suffit de considérer la grammaire définie par les règles

$$A \rightarrow a ; S \rightarrow ab + aS$$

avec pour axiomes  $A$  et  $S$ .

*5.* De même, on peut ajouter le mot  $\epsilon$  au langage du troisième exemple en ajoutant à la grammaire un axiome  $E$  et la règle  $E \rightarrow 1$ , ce qui conserve la propriété *NTS*.

Le théorème suivant généralise ces exemples.

**Théorème 8.** *Les langages rationnels sont NTS.*

*Démonstration.* Soit  $L$  un langage rationnel sur l'alphabet  $A$  et  $M$  son monoïde syntaxique. Considérons la grammaire qui a pour variables les éléments de  $M$ , et qui a pour règles toutes les règles de la forme  $m_1 \rightarrow m_2 m_3$  où  $m_1, m_2, m_3 \in M$  et  $m_1 = m_2 m_3$ , auxquelles on ajoute les règles  $m \rightarrow x$  si  $x$  est une lettre terminale de la classe  $m$ , et la règle  $E \rightarrow \epsilon$  où  $E$  est l'élément neutre du monoïde

syntaxique. Considérons un morphisme de monoïdes  $\phi : (A \cup M)^* \rightarrow A^*$  tel que  $\phi(a) = a$  pour chaque  $a \in A$ , et pour chaque  $m \in M$ ,  $\phi(m)$  soit dans la classe  $m$ . On voit alors que pour chaque  $m \in M$ ,  $\widehat{L}_G(m)$  est l'image inverse par  $\phi$  de la classe  $m$ . De plus pour chaque règle  $u \rightarrow v$  de la grammaire,  $\phi(u)$  et  $\phi(v)$  sont dans la même classe syntaxique. Cette grammaire est donc NTS.

Il suffit donc de choisir pour axiomes les éléments de  $M$  composés de mots de  $L$  pour obtenir une grammaire reconnaissant  $L$ .  $\square$

D'autre part, les langages de Dyck sont NTS. En effet, le langage  $D_n^*$  est engendré par la grammaire NTS

$$S \rightarrow SS + \sum_{i=1}^n a_i S \bar{a}_i + 1.$$

Le langage  $D_n$  est aussi NTS car il est engendré par la grammaire

$$\left\{ \begin{array}{l} T \rightarrow \sum_{i=1}^n a_i S \bar{a}_i \\ S \rightarrow SS + \sum_{i=1}^n a_i S \bar{a}_i + 1. \end{array} \right.$$

La propriété de clôture la plus importante vérifiée par cette classe de langages est la clôture par intersection avec les langages rationnels [3].

Une étude plus complète des propriétés des langages NTS peut être trouvée dans [3] et [2]. En particulier, ces langages sont reconnaissables par un automate à pile déterministe (ils constituent donc une sous-classe stricte des langages algébriques) ; l'équivalence de deux grammaires NTS est décidable ainsi que le problème consistant à dire si une grammaire donnée est NTS [2]. Par contre, le problème d'inclusion des grammaires NTS est indécidable [2].

## Références

- [1] J. Bestel and L. Boasson. Context-free languages. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 2, pages 59–102. Elsevier, 1990.
- [2] G.Senizergues. The equivalence and inclusion problems for NTS languages. *Journal of Computer and System Sciences*, 31 :303–331, 1985.
- [3] L.Boasson and G.Senizergues. NTS languages are deterministic and congruential. *Journal of Computer and System Sciences*, 31 :332–342, 1985.