

L'IDEE DE GRAMMAIRE AVEC LE CONTEXTE NATUREL

Leszek Haduch

Institute of Informatics

Technical University of Lodz

Lodz, ul. Piotrkowska 220, Poland

ABSTRACT

Commonly used grammars which describe natural languages /ex. ATN, Metamorphosis Grammars/ can be hardly applied in describing highly inflectional languages. So I propose a grammar called the grammar with natural context which takes into consideration properties of highly inflectional languages /ex. Polish / as well as structural languages /ex. English/. I introduce its normal form.

1. Introduction

Des grammaires introduits par Woods¹ /ATN/ et par Colmerauer² / Grammaire Metamorphique/ ont trouvé une large application pour l'analyse et le synthèse des langages naturels. Elles ont été appliqués aussi pour l'analyse des phrases écrites du polonais /Szpakowicz³, Maligórski⁴/.

Toutefois, de ces derniers travaux on peut conclure, que:

- application n'importe quelle de susdites grammaires occasionne une formalisation du langage.
- ces grammaires ne consentent pas à

l'entrelacement des groupes, qui apparaît souvent aux langages flexionnels.

- assez souvent la description syntaxique d'un groupe c'est une permutation d'un certain nombre des unités syntaxique, que suggère une idée de l'introduction d'une dynamique à la description statique de la grammaire.

Prenant ces conclusions en considération dans la suivante partie de l'article je propose une certaine grammaire, que j'appelle la grammaire avec le contexte naturel. Elle consent à l'entrelacement des groupes et aussi introduit une dynamique à la description statique de la grammaire.

2. Conception de la grammaire avec le contexte naturel et sa forme normale.

Définition 1.1

La grammaire avec le contexte naturel est 5-uplet:

$$G = \langle V, \Sigma, \mathcal{H}, P, B \rangle$$

où:

V - l'ensemble fini des symboles

Σ - son sous-ensemble fini /l'ensemble des symboles terminaux/

\mathcal{K} - l'ensemble fini des noyaux du contexte / $\mathcal{K} \subset V^+$ /

S- le symbole / $S \in V - \Sigma$ /

P- l'ensemble fini des règles de la grammaire à la forme suivante:

$$x_1, x_2, \dots, x_n, \omega, z_1, z_2, \dots, z_m \rightarrow A$$

où:

ω - un noyau du contexte - $\omega \in V^+ \wedge \omega \in \mathcal{K}$

A - un symbole de l'ensemble $V - \Sigma$

x_1, x_2, \dots, x_n - le gauche contexte, $x_i \in V^*$ pour $i \in [1, n]$

z_1, z_2, \dots, z_m - le droit contexte, $z_j \in V^*$ pour $j \in [1, m]$

Acceptons hypothèse, que pour quelconque

$k, l \leq |\mathcal{K}|$, $\omega_k \neq \omega_l$ pour $k \neq l$

Introduisons les désignations nécessaires pour la description des langages engendré par la grammaire:

Soit

$$x_1, x_2, \dots, x_n, \omega, z_1, z_2, \dots, z_m \rightarrow A \in P \text{ et } y, t \in V^*$$

Soit

$$y = y_1 x_1 y_2 x_2 \dots y_n x_n y_{n+1} \text{ avec } y_i \in V^*, i \in [1, n+1]$$

soit

$$t = t_1 z_1 t_2 z_2 \dots t_m z_m t_{m+1} \text{ avec } t_j \in V^*, j \in [1, m+1]$$

alors

$$y \omega t \Rightarrow y' A t' \text{ où: } y' = y_1 y_2 \dots y_{n+1} \\ t' = t_1 t_2 \dots t_{m+1}$$

Le langage engendré par la grammaire

avec le contexte naturel est l'ensemble:

$$L(G) = \{ \omega : \omega \in \Sigma^* \wedge \omega \Rightarrow^* S \}$$

Passons au exemple de la grammaire avec le contexte naturel:

L'ensemble $V = \{ S, a, b, c \}$

$$\Sigma = \{ a, b, c \}$$

$$\mathcal{K} = \{ SS, b \}$$

$$S = S$$

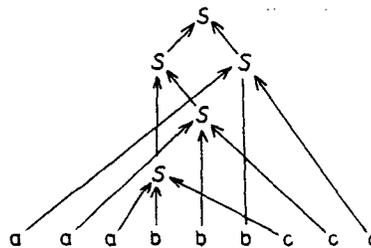
L'ensemble P se compose des règles suivantes:

$$S \leftarrow SS$$

$$S \leftarrow abc$$

La grammaire susdite décrit aussi bien série $a^n b^n c^n$, par exemple, pour $n=3$

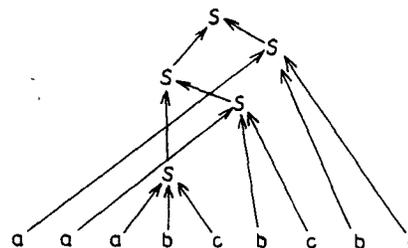
l'arbre de la déduction sera suivant:



et aussi:

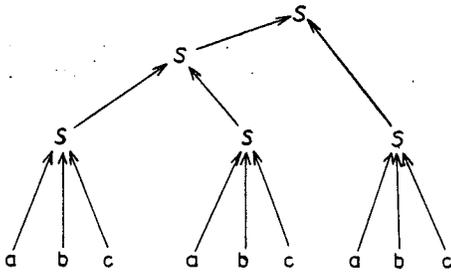
$$a^n (bc)^n$$

par exemple, pour $n=3$:



et:

$(abc)^n$ par exemple, pour $n=3$



etc. Ainsi donc cette grammaire décrit toute la famille des langages à contexte libre et des langages contextuels.

Définition 1.2

La forme normale de la grammaire avec le contexte naturel est une telle grammaire $G = \langle V, \Sigma, \mathcal{H}, P, S \rangle$ pour laquelle:

1/ $n, m = 1$

2/ $|\omega| < 2$ pour toutes les règles

$$x_1, x_2, \dots, x_n, \omega, z_1, z_2, \dots, z_m \rightarrow A \in P$$

Proposition 1.1

Pour toutes les grammaires avec le contexte naturel $G = \langle V, \Sigma, \mathcal{H}, P, S \rangle$ on peut construire une certaine grammaire avec le contexte naturel à la forme normal $G' = \langle V', \Sigma, \mathcal{H}', P', S \rangle$, que engendre le même langage.

Démonstration:

Soit.

$$x_1, x_2, \dots, x_n, \omega, z_1, z_2, \dots, z_m \rightarrow A \in P$$

Prenons en considération des cas suivants:

1/ $n=1$ et $m=1$, cette règle participe à P'

2/ $n > 1$ et $m > 1$, alors l'ensemble des règles P' nous construisons à la manière suivante:

$$\begin{aligned} x_n, \omega, z_1 &\rightarrow D_1 \\ x_{n-1}, D_1, z_2 &\rightarrow D_2 \\ x_{n-2}, D_2, z_3 &\rightarrow D_3 \\ \vdots &\vdots \\ x_{n-m+1}, D_{m-1}, z_m &\rightarrow D_m \\ x_{n-m}, D_m, \varepsilon &\rightarrow D_{m+1} \\ \vdots &\vdots \\ x_1, D_{n-1}, \varepsilon &\rightarrow A \end{aligned}$$

si $m > n$:

$$\begin{aligned} x_n, \omega, z_1 &\rightarrow E_1 \\ x_{n-1}, E_1, z_2 &\rightarrow E_2 \\ \vdots &\vdots \\ x_1, E_{n-1}, z_{m-n} &\rightarrow E_n \\ \varepsilon, E_n, z_{m-n-1} &\rightarrow E_{n+1} \\ \vdots &\vdots \\ \varepsilon, E_{m-1}, z_m &\rightarrow A \end{aligned}$$

3/ $n=1$ et $m > 1$, alors l'ensemble des règles P' nous construisons à la manière suivante:

$$\begin{aligned} x_1, \omega, z_1 &\rightarrow F_1 \\ \varepsilon, F_1, z_2 &\rightarrow F_2 \\ \vdots &\vdots \\ \varepsilon, F_{m-1}, z_m &\rightarrow A \end{aligned}$$

4/ $n > 1$ et $m=1$, alors

$$\begin{aligned} x_n, \omega, z_1 &\rightarrow C_1 \\ x_{n-1}, C_1, \varepsilon &\rightarrow C_2 \\ \vdots &\vdots \\ x_1, C_{n-1}, \varepsilon &\rightarrow A \end{aligned}$$

où: C_i, D_i, E_j, F_j - des neufs symbols,
 $i \in [1, n-1]$ et $j \in [1, m-1]$

Considérons la deuxième propriété de la grammaire avec le contexte naturel à la forme normale:

- si $|\omega| < 2$ et il est réalisé la première propriété, cette règle nous ajoutons à l'ensemble P' .
- si $|\omega| > 2$, $\omega = \omega_1 \omega_2 \dots \omega_l$ où: $l > 2$ et $\omega_1 \in V$ et aussi il est réalisé la première propriété, alors l'ensemble P' nous construisons à la manière suivante:

$$\begin{aligned} x_1, x_2, \dots, x_n, \omega_1, \varepsilon &\longrightarrow B_1 \\ \varepsilon, B_1 \omega_2, \varepsilon &\longrightarrow B_2 \\ \varepsilon, B_2 \omega_3, \varepsilon &\longrightarrow B_3 \\ \varepsilon, B_3 \omega_4, \varepsilon &\longrightarrow B_4 \\ \vdots &\vdots \\ \varepsilon, B_{l-1} \omega_l, z_1, z_2, \dots, z_m &\longrightarrow A \end{aligned}$$

où: B_k - un neuf symbole, $k \in [1, l-1]$

L'ensemble V' sera l'ensemble V augmenté par des nouveaux symboles venant pendant les constructions susdites. L'ensemble sera l'ensemble aussi augmenté par des nouveaux symboles venant pendant les constructions susdites. Par construction, la grammaire $G' = \langle V', \Sigma; \mathcal{K}, P', S \rangle$ est sous la forme normale et il est clair que $L(G') = L(G)$

3. Conclusion

La grammaire susdite a été appliqué à la description syntactique du polonais et du français. Au cas du polonais, il

s'est montré être possible seulement l'entrelacement du gauche contexte par l'introduction une limitation du cercle de l'entrelacement c'est-à-dire pour ne sortir pas des phrases subordonnées à la phrase principale. Au cas du droit contexte il est assez difficile trouver le symbole identifiant la fin de la phrase subordonnée. Par suite de cela, existant un système de l'analyse automatique du polonais écrit n'admet pas à l'entrelacement du droit contexte. En l'algorithme de l'analyse j'ai exploité une certaine modification de l'algorithme LR/k/. C'est un algorithme intermédiaire entre un classique algorithme "bottom-to-up" et l'algorithme LR/k/. L'application pratique de la grammaire avec le contexte naturel mène aux conclusions suivantes:

- au contraire des grammaires de métamorphoses et d'ATN à la grammaire avec le contexte naturel il peut avoir lieu l'entrelacement des groupes
- on peut appliquer des méthodes de l'accélération de l'analyse en modifiant des connues méthodes des grammaires formales.
- il a été introduit une certaine dynamique à la statique description de la grammaire

REFERENCES

- [1] Woods W.A., Transition network grammars for natural language analysis,

C.ACM 13, 1970.

- [2] Colmerauer A., Metamorphosis Grammars. In Bolc L./ed/ Natural Language Communication with Computers, Lecture Notes in Computer Science 63, 1978.
- [3] Szpakowicz S., Automatyczna analiza składniowa polskich zdań pisanych, Ph.D. thesis, 1978.
- [4] Waligórski S. et al., Projekt i oprogramowanie modułu analizy języka naturalnego Design and Implementation of the Natural Language Analysis Module . Vol. I, 1979, Vol. II, 1980, Institute of Informatics, Warsaw University.
- [5] Greibach S., Hopcroft J., Scattered context grammars, J. Comput. Syst. Sci., 1969, 3, 233-247.