APPRENTISSAGE DE GRAMMAIRES CATEGORIELLES DE MODULES

Travail de recherche fait par:

Camilo Thorne

Sous la direction de:

M. D. BECHET (LIPN)

Villetaneuse, le 1er octobre 2004.

Structure de la présentation

- Notions générales.
- Grammaires de modules étiquetées.
- Résultats: apprentissage de grammaires de modules étiquetées.

Définition – grammaires catégorielles

- Une <u>grammaire catégorielle</u> est un triplet

$\mathbf{G} = < \mathbf{\Sigma} \ , \boldsymbol{\delta} \ , \mathbf{S} >$ avec :

- $\mathbf{\Sigma}$ un alphabet fini (ou lexique).
- $\boldsymbol{\delta} : \mathbf{\Sigma} \rightarrow \mathbf{P_f(T)}$ une fonction finie.
- **S** le type des phrases bien formées.

Caractères généraux

- Leur système de types, **LC** (ou **MINCLL**), relève de la logique linéaire.
- C’est un formalisme faiblement équivalent à celui des grammaires hors-contexte (Kanazawa, 1998).
- L’analyse syntaxique se fait via un arbre ou réseau de preuve pour **MINCLL**.

Le langage engendré par une grammaire catégorielle

$$s = m_1 \ldots m_n \in \mathbf{L(G)}$$

$$\Leftrightarrow_{df}$$

• ils existent n types $A_1, \ldots, A_n$ assignés par δ tels que le séquent $A_1, \ldots, A_n \vdash S$ est démontrable dans **MINCLL**

Apprentissage – conditions suffisantes

Densité finie bornée

$\Rightarrow$ (1)

Élasticité finie

$\Rightarrow$ (2)

Apprenabilité

(1) Shinohara, 1990.

(2) Kanazawa, 1998.

Densité finie bornée

- Cet une propriété des systèmes grammaticaux **GS** = <**S***,**GK**, **L**> associés à une classe **GK** de grammaires.

- Elle implique, pour chaque $n \in N$, que la classe $\{$**L(G)** | **G**$\in$ **GK**, t(**G**)$\leq$ n et **G** est réduite par rapport à $X \subseteq$ **S*** fini$\}$ a une élasticité finie.

Élasticité finie

- Cet une propriété d’une classe ***L*** de langages sur un alphabet **Σ** .

- _Quand on fait le lien avec une classe **GK** de grammaires, elle implique l’existence d’un algorithme d’apprentissage φ_{GK} pour **GK**.

Apprenabilité

Une classe GK est apprenable lorsqu’il existe un algorithme φ_{GK} d’apprentissage pour cette classe.

Grammaires de modules étiquetées

- Ce sont des grammaires ayant un alphabet « enrichi » avec des informations syntaxiques: des modules. Leur alphabet est donc une partie de $\mathbf{\Sigma} \times \mathbf{M}$.

- Si on prend pour alphabet $\mathbf{\Sigma}_{\mathbf{mod}} = \mathbf{\Sigma} \times \mathbf{M}$ on a affaire à $\mathbf{GC}_{\mathbf{mod}}$, aux grammaires étiquetées d'alphabet infini.

- Si on prend comme alphabet $\mathbf{\Sigma}_{\mathbf{modf}} \subset \mathbf{\Sigma} \times \mathbf{M}$ finie on a affaire à $\mathbf{GC}_{\mathbf{modf}}$, aux grammaires étiquetées d'alphabet fini.

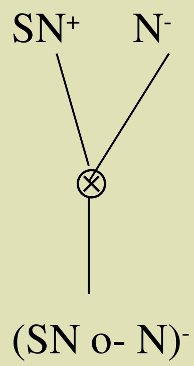
Un exemple

- Voici une grammaire grammaire étiquetée d’alphabet fini (défini implicitement par δ):
- Soit $\mathbf{G}_0$ la grammaire de modules avec δ définie par:

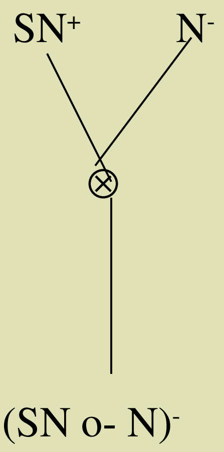
$N^{-\bullet}$

chat $\quad\mapsto\quad N^{-\bullet}$

Un exemple



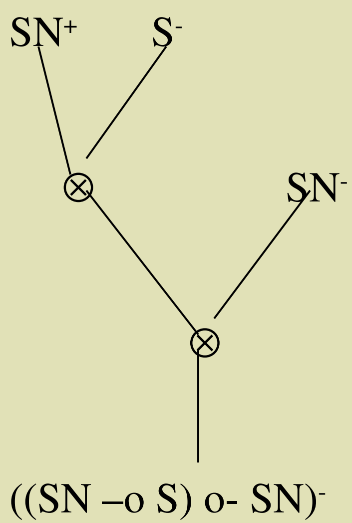
le $\mapsto$



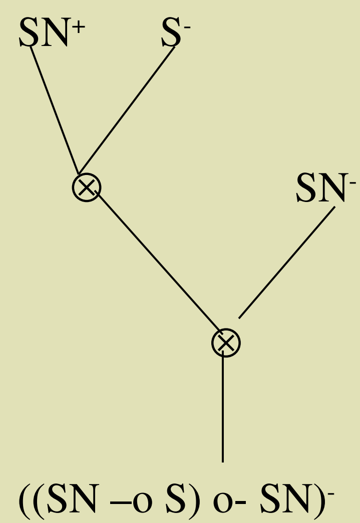
$N^{-\bullet}$

escalier $\mapsto$ $N^{-\bullet}$

Un exemple



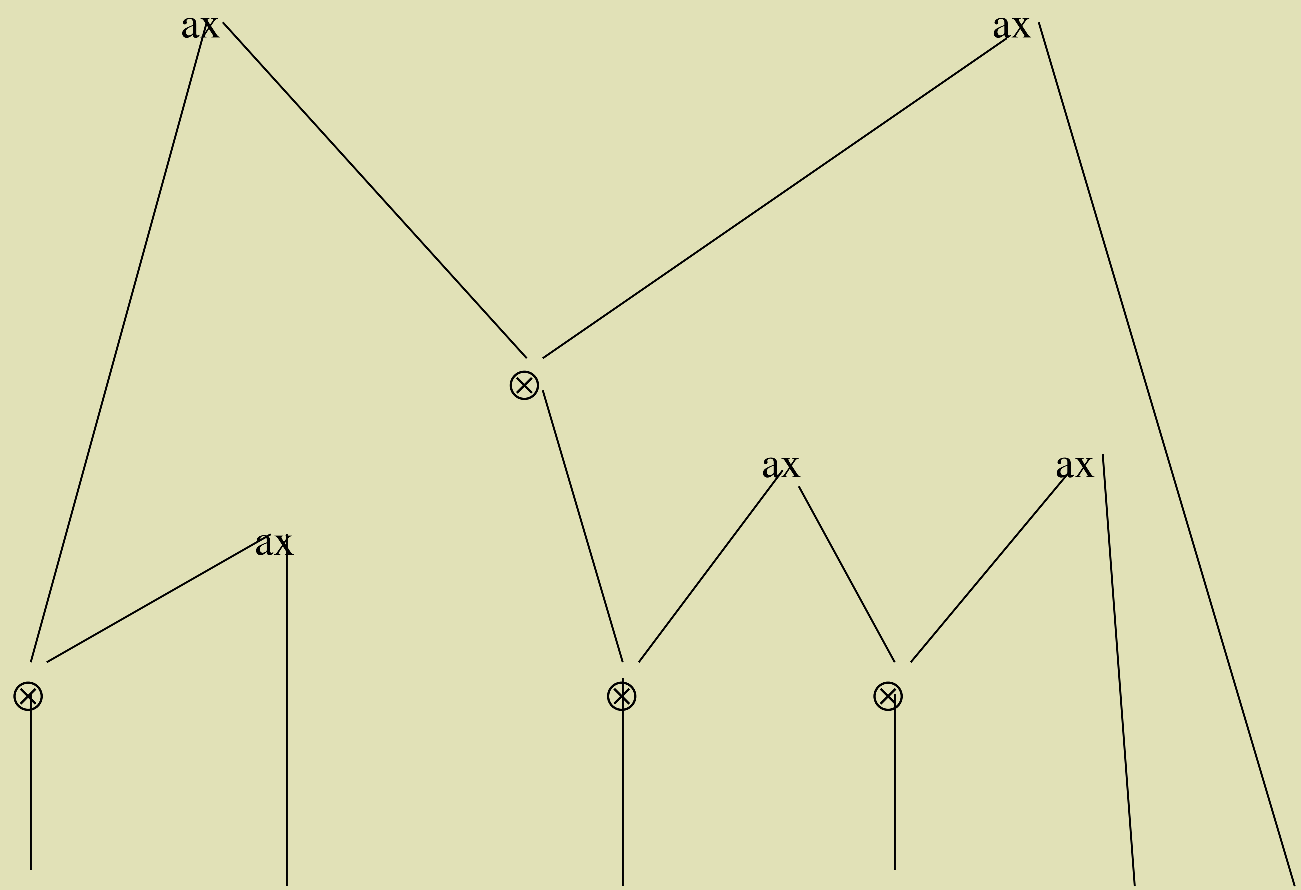
monte $\mapsto$



- On peut montrer que « Le chat monte l'escalier » $\in$ **L(G$_0$).**

- On obtient un réseau de preuve planaire:

Un exemple



|— (SN o- N)$^-$, N$^-$, ((SN –o S) o- SN)$^-$, (SN o- N)$^-$, N$^-$, S$^+$

Résultats

➢ $L(\mathbf{GC}_{\mathbf{modf}})$ a une élasticité finie. Donc $\mathbf{GC}_{\mathbf{modf}}$ est apprenable.

➢ Ils existent des sous-classes apprenables de $\mathbf{GC}_{\mathbf{mod}}$, car le système $\mathbf{GS}_{\mathbf{mod}}$ a une densité finie bornée.

➢ $L(\mathbf{GC}_{\mathbf{mod}})$ a une élasticité infinie.

Quelques références

- Kanzawa. Learnable classes of Categorial Grammars. 1998.
- Béchet. Incremental Parsing of Lambek Calculus Using Proof-Net Interphases. 2002.
- Shinohara. Inductive Inference from Positive Data is Powerful. 1990.
- Lambek. The Mathematics of Sentence Structure. 1958.
- Tiede. Deductive Systems and Grammars. Proofs as Grammatical Structures. 1999.

Théorèmes cités

- **THEO**. Soit **GK** une classe de grammaires. Si ***L*(GK)** a une élasticité finie, alors **GK** est apprenable. (Kanazawa, 1998).

- **THEO**. Soit **GS** un système grammatical. Si **GS** a une épaisseur finie bornée, alors la classe {**L(G)** | **G** ∈ **GK**, t(**G**)≤ n et **G** est réduite par rapport à *X*} a une élasticité finie. (Shinohara, 1990).

- **THEO**. Un séquent $\Gamma \vdash A$ est démontrable en **MINCLL** avec contexte vide éventuellement vide ⇔ il existe un réseaux de preuve planaire pour $\vdash \Gamma^{-}, A^{+}$. (Retoré, 1996).

Annexe 1: apprentissage

Définition – apprentissage

- Une classe **GK** de grammaires est <u>apprenable à partir d'exemples positifs</u>

$\Leftrightarrow_{df}$

Il existe une algorithme d'apprentissage φ_{GK}, i.e. un algorithme qui converge à terme vers une certaine grammaire $\mathbf{G} \in \mathbf{GK}$ en ayant comme entrée l'énumération d'un $\mathbf{L} \in \mathit{L}(\mathbf{GK})$ quelconque.

Définition - algorithme d'apprentissage

- Soit **GK** une classe de grammaires. On dit que l'algorithme φ_{GK} <u>apprend</u> **GK**

$\Leftrightarrow_{df}$

- $\varphi_{GK} : \cup_{i \in N} (S^*)^i \rightarrow GK$

- pour toute grammaire **G** avec **L(G)** = $<s_i>_{i \in N}$, il existent une grammaire **G'** et un entier n_0 tel que pour tout $n > n_0$, $\varphi_{GK}(<s_0,\ldots s_{n0}>)$ = **G'** et **L(G) = L(G')**

Définition – élasticité finie

- Une classe $\boldsymbol{L}$ de langages sur un alphabet $\boldsymbol{\Sigma}$ a une <u>élasticité finie</u>

$\Leftrightarrow_{df}$

il n'existe pas de suites $<\mathbf{L}_i>_{i\in N}$, $<\mathbf{s}_i>_{i\in N}$ de langages et de chaînes sur $\boldsymbol{\Sigma}$ telles que, pour tout $i\in N$:

- $\mathbf{s}_i \notin \mathbf{L}_i$
- $\{\mathbf{s}_0,\ldots,\mathbf{s}_i\} \subseteq \mathbf{L}_{i+1}$

Définition - systèmes grammaticaux

➢ Un <u>système grammatical</u> est un triplet

GS = < **S***, **GK**, **L** > avec:

- **S*** $\subseteq$ **Σ** * un ensemble récursif de suites sur un alphabet **Σ** .
- **GK** une classe de grammaires.
- **L : GK → P(S*)** telle que, pour toute **G** $\in$ **GK**, **L(GK)** $\subseteq$ **S***.

Définition – grammaire réduite par rapport à X

➢Soit **GS** un système grammatical. $\mathbf{G} \in \mathbf{GK}$ et $X \subseteq \mathbf{S}^*$ fini. **G** est <u>réduite par rapport à X</u>

$\Leftrightarrow_{df}$

- $X \subseteq \mathbf{L}(\mathbf{G})$
- pour toute $\mathbf{G'} \in \mathbf{GK}$, si $\mathbf{G'}\ R\ \mathbf{G}$ et $\mathbf{G'} \neq \mathbf{G}$, alors

$\neg[X \subseteq \mathbf{L}(\mathbf{G'})]$

Définition – densité finie bornée

➢ Un système grammatical **GS** a une <u>densité finie bornée</u>

$\Leftrightarrow_{df}$

• pour tout i,j∈ N, $\mathbf{G}_i$ *R* $\mathbf{G}_j \Rightarrow \mathbf{L}(\mathbf{G}_i) \subseteq \mathbf{L}(\mathbf{G}_j)$ (**L** est croissante, *R* est l'ordre des grammaires)

• pour tout n∈ N, la classe {**L(G)** | **G**∈ **GK**, t(**G**)≤ n et **G** est réduite par rapport à $X \subseteq$ **S*** fini} est une classe finie.

Annexe 2: types

Grammaires de modules

➢Ce sont des grammaires où l'on remplace l'ensemble **T** des types par **M** celui des modules de la logique linéaire. En effet, il se trouve que:

Existence d'un arbre d'analyse

$\Leftrightarrow$ (*)

Existence d'un réseau de preuve planaire

(*) Retoré, 1996.

Le système de types

- L'ensemble des **<u>types</u>** est défini par la grammaire :

$$T ::= At \mid T \multimap T \mid T \circ\!- T \mid T \otimes T.$$

- Comme **<u>système de types</u>** on a un calcul de séquents, la logique linéaire multiplicative non commutative bilatère sans constantes ni opérateur $\wp$. **MINCLL**, pour abréger.