Arbres et discontinuité

1 Introduction

2 Les arbres

2.1 Définition à base de graphes

2.1.1 Graphes

Un graphe est donné par un couple $\langle X, \Gamma \rangle$ où X est un ensemble et Γ une application de X dans $\mathcal{P}(X)$.

Tout élément x de X est un *sommet* du graphe (ou noeud).

$\Gamma(x)$ est l'ensemble des *successeurs* ou des *fils* ou des *descendants immédiats* de x.

Deux sommets x et y qui sont tels que y appartient à $\Gamma(x)$ forment un *arc*; x est l'*origine* de l'arc, y en est l'*extrémité*.

$\Gamma^*(x)$ est l'ensemble des *descendants* de x : il comprend x, les fils de x, les fils des fils de x, etc. :

(i) $x \in \Gamma^*(x)$
(ii) $y \in \Gamma(x)$ et $z \in \Gamma^*(y) \Rightarrow z \in \Gamma^*(x)$

2.1.2 Chaînes, cycles

Une *chaîne* est constituée par une suite $\langle x_1, x_2, \ldots, x_q, x_{q+1}\rangle$ de sommets du graphe vérifiant:

$$\forall i \leq q \;\; x_i \in \Gamma(x_{i+1}) \text{ ou } x_{i+1} \in \Gamma(x_i)$$

x_1 et x_{q+1} sont alors les deux extrémités de la chaîne.

Un *cycle* est constitué par une chaîne dont les deux extrémités coïncident.

Un graphe est *connexe* si et seulement si pour tout x et pour tout y de X (x et y étant distincts) il existe une chaîne admettant pour extrémités x et y.

2.1.3 Graphes étiquetés

Dans un graphe, tous les sommets sont différents. On peut affecter des *étiquettes*, prises dans un ensemble E, aux sommets.

Un graphe est à ce moment-là défini par $\langle X, \Gamma, L \rangle$ où L est une application de X dans l'ensemble E.

Les arcs peuvent aussi être étiquetés.

2.1.4 Première définition des arbres

Un arbre est un graphe connexe et sans cycles.

Attention, cette définition n'est pas celle que l'on prend en général en linguistique (ou en informatique): elle ne suppose pas l'existence d'une racine unique.

2.1.5 Deuxième défintion

Une *arborescence* est un arbre $A = \langle X, \Gamma \rangle$ qui comporte un sommet a tel que tout sommet $x \in X$ est un descendant de a.

Ce sont les arborescences que l'on appelle généralement arbres.

2.1.6 Définitions et propriétés

Etant donné un arbre $A = \langle X, \Gamma \rangle$, $\Gamma(x)$ est l'ensemble des *fils* du sommet x.

Un sommet est le successeur (ou le fils) au maximum d'un autre sommet.

Il n'y a qu'un sommet qui n'est fils d'aucun autre sommet: la *racine* de l'arbre.

Une *feuille* est un sommet qui n'a aucun fils.

Si $y \in \Gamma^*(x)$, on dit que x *domine* y.

2.1.7 Arbres ordonnés

$A = \langle X, \delta \rangle$ est un *arbre ordonné* à condition que δ soit une application de X dans X^* telle que, pour tout x, si $\delta(x) = x_1 x_2 \ldots x_p$, tous les x_i sont différents et si $\Delta(x)$ est l'ensemble composé de tous les x_i, alors $\langle X, \Delta \rangle$ vérifie les propriétés des arborescences.

On peut définir sur un arbre ordonné $A = \langle X, \delta \rangle$ un *ordre de précédence* (ordre strict partiel), de la manière suivante:

(i) si $\delta(x) = x_1 x_2 \ldots x_p$ et $i < j$ alors x_i précède x_j

(ii) si u précède v, u domine x et v domine y, alors x précède y

L'ensemble des feuilles d'un arbre ordonné est totalement ordonné par la relation de précédence. On peut par conséquent écrire:

$$feuilles(A) = z_1 z_2 \ldots z_n$$

2.2 Définition alternative des arbres ordonnés étiquetés (Wall, 1972)

Un arbre est un quadruplet $A = \langle X, L, D, P\rangle$ où X est un ensemble, L une application de X dans un ensemble Cat, D est une relation d'ordre partiel sur X (relation de domination) et P est une relation d'ordre strict partiel sur X (relation de pécédence). Il faut que les trois conditions suivantes (racine unique, exclusivité et non-croisement) soient vérifiées :

(i) $\exists x \in X \; \forall y \in X \; \langle x, y\rangle \in D$

(ii) $\forall x, y \in X \; \langle x, y\rangle \in P$ ou $\langle y, x\rangle \in P \Leftrightarrow \langle x, y\rangle \notin D$ et $\langle y, x\rangle \notin D$

(iii) $\forall w, x, y, z \in X$

$$\langle w, x\rangle \in P \text{ et } \langle w, y\rangle \in D \text{ et } \langle x, z\rangle \in D \Rightarrow \langle y, z\rangle \in P$$

2.2.1 Représentation d'un énoncé par un arbre

On considère un ensemble W, l'ensemble des formes lexicales, et on suppose qu'il existe une application:

$$cat : W \to \mathcal{P}(Cat)$$

où Cat est un ensemble de catégories.

Soit A un arbre tel que $feuilles(A) = z_1 z_2 \ldots z_n$ et soit $w_1 w_2 \ldots w_n \in W^*$ un énoncé. Cet énoncé sera représenté par l'arbre A si et seulement si:

$$\forall i \in \{1, \ldots, n\} \;\; L(z_i) \in cat(w_i)$$

2.3 Arbres et grammaires

2.3.1 Définition des grammaires indépendantes du contexte

Une grammaire *indépendante du contexte* (ou “context free”, ou CF) est décrite par un 4-uple $G = \langle V_T, V_N, S, \mathcal{R} \rangle$ où V_T est un vocabulaire *terminal* (ou alphabet), V_N est un vocabulaire auxiliaire (ou vocabulaire *non terminal*) disjoint de V_T, S est un élément particulier de V_N, appelé *axiome* de la grammaire, et $\mathcal{R}$ est un ensemble de *règles* (de réécriture) qui ont la forme suivante:

$$A \to \beta$$

A étant un élément de V_N et β étant un élément de $(V_N \cup V_T)^*$.

Dans une règle telle que $A \to \beta$, A est la *partie gauche* et β est la *partie droite*.

2.3.2 Décomposition d'un arbre en arbres élémentaires

Etant donné un arbre ordonné étiqueté $\langle X, \delta, L\rangle$, pour chaque sommet $x \in X$ qui n'est pas une feuille, on définit l'*arbre élémentaire* de racine x $A[x] = \{X_x, \delta_x, L_x\}$ où :

$$X_x = \{x\} \cup \Delta(x)$$
$$\delta_x(x) = \delta(x)$$
$$\forall y \in \Delta(x) \;\; \delta_x(y) = \varepsilon$$
$$\forall y \in X_x \;\; L_x(y) = L(y)$$

2.3.3 Arbres élémentaires et règles de grammaires

On considère à présent des arbres dont les sommets sont étiquetés par des valeurs prises dans $V_N \cup V_T$.

Etant donnée une grammaire CF $G = \langle V_T, V_N, S, \mathcal{R} \rangle$, on dira qu'un arbre élémentaire $A[x] = \{X_x, \delta_x, L_x\}$ tel que $\delta(x) = x_1 \ldots x_n$ est compatible avec la grammaire si et seulement si

$$L(x) \to L(x_1) \ldots L(x_n)$$

est une règle de la grammaire.

On dira qu'un arbre A est compatible avec la grammaire, ou qu'il est *admissible* relativement à la grammaire, si et seulement si, pour tous les sommets x qui ne sont pas des feuilles, $A[x]$ est compatible avec la grammaire.

2.4 Les représentations en machine des arbres

2.4.1 Représentation par des tableaux à deux dimensions

1°/ Le tableau des fils de chaque sommet.
2°/ Les relations de dominance et de précédence.

2.4.2 Représentation par le tableau des fils aînés et le tableau des frères cadets

2.4.3 Représentation dans des tableaux aux dimensions variables

2.4.4 Représentations dans les langages de manipulation de symboles

2.5 Arbres et structures parenthésées