

Interface syntaxe - linguistique computationnelle
Examen final
Aucun document autorisé ; Durée : 1 heure 30.

1. Soit la définition suivante d'une formule logique (langage L_p) :

- (i) Si A est un nom de prédicat du vocabulaire de L_p , et chacun des $t_1...t_n$ une constante ou une variable du vocabulaire de L_p , alors $A(t_1, ..., t_n)$ est une formule.
- (ii) Si φ est une formule dans L_p , alors $\neg\varphi$ l'est aussi.
- (iii) Si φ et ψ sont des formules de L_p , alors $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$ sont des formules de L_p .
- (iv) Si φ est une formule et x une variable, alors $\forall x\varphi$ et $\exists x\varphi$ sont des formules de L_p .
- (v) Rien d'autre n'est une formule

- (a) En fixant arbitrairement un vocabulaire **fini** pour L_p (noms de variables, noms de constantes et noms de prédicats), et en supposant pour simplifier que tous les prédicats sont unaires, proposer une grammaire hors-contexte (CFG) qui reconnaît ce langage.
- (b) Donner les arbres syntaxiques correspondants aux expressions (en supposant que les constantes non logiques appartiennent au vocabulaire) $(P(x) \rightarrow Q(a))$ et $\forall x\exists y((F(x) \wedge A(y)) \rightarrow B(x))$
- (c) Pour supprimer la contrainte de finitude sur les constantes non logiques, on pourrait imaginer d'autoriser des expressions rationnelles en partie droite des règles de réécriture. Par exemple, les règles "terminales" pour les constantes individuelles ($C \rightarrow a \mid b \mid c$) seraient remplacées par $C \rightarrow c(0|1)^*$ (une constante individuelle se réécrit sous la forme d'un c suivi d'un nombre quelconque de 0 ou de 1), ce qui permet d'avoir un ensemble infini (dénombrable) de constantes individuelles. Il faut bien sûr redéfinir la notion de dérivation en conséquence. Une telle extension augmente-elle le pouvoir expressif des grammaires algébriques ? Justifier.
- (d) Supposons que l'on s'autorise à supprimer les parenthèses rendues inutiles par l'associativité de $\wedge$ et $\vee$. (Ainsi, $((\alpha \wedge \beta) \wedge \gamma)$ s'écrit $(\alpha \wedge \beta \wedge \gamma)$). Modifier la grammaire précédente pour reconnaître ce nouveau langage. Vérifier le bon fonctionnement de la grammaire (qui peut être ambiguë) sur l'expression $(\alpha \wedge \beta \wedge (\alpha \vee \gamma) \wedge (\alpha \vee \beta))$.
- (e) Toujours pour simplifier l'écriture, on veut autoriser cette fois la suppression des parenthèses dans des expressions combinant les connecteurs $\wedge$ et $\vee$: $\alpha \wedge \beta \vee \gamma$ doit être reconnu. Proposer une grammaire *non ambiguë* reconnaissant ce langage, en supposant une *priorité* entre les connecteurs : on posera que $\wedge$ est prioritaire par rapport à $\vee$, ce qui a pour conséquence que $\alpha \wedge \beta \vee \gamma$ est interprétée comme $((\alpha \wedge \beta) \vee \gamma)$. (Indice : dissymétriser les opérandes.)

2. Soit la grammaire

S	→	NP VP	D	→	un le
NP	→	D N	N	→	chat chien
		PN	PN	→	Jean Marie
VP	→	V_i	V_i	→	dort ronfle
		V_t NP	V_t	→	voit connaît
		V_a CP	V_a	→	pense croit
CP	→	que S			

- (a) Donner l'arbre syntaxique de la phrase (1).
 - (1) Jean croit que le chat dort
- (b) Quel coût représenterait pour une analyse syntaxique ascendante l'ajout à cette grammaire de la règle " $V_i \rightarrow$ pense" ? Illustrer la réponse en représentant (la partie pertinente de) l'analyse syntaxique ascendante (*shift/reduce*) de (2).
 - (2) Jean pense que Marie pense que Jean dort
- (c) On suppose que la phrase (3a) peut recevoir une représentation logique de la forme (3b).
 - (3) a. Jean croit que Marie dort.
 - b. $\text{croit}(j, \text{dort}(m))$ ou, avec une notation à la λ -calcul : $((\text{croit})j)(\text{dort})m$
 Donner le détail des λ -expressions "lexicales" et des règles de composition nécessaires pour construire compositionnellement cette représentation logique avec la grammaire précédente.
- (d) Comment traiter (d'une façon simple, à défaut d'être linguistiquement plausible) le cas de (4) ?
 - (4) a. Jean pense voir Marie
 - b. $\text{pense}(j, \text{voir}(j, m))$ ou $((\text{pense})j)((\text{voir})j)m$

On envisagera deux pistes pour traiter ce cas : soit essentiellement par la syntaxe, en supposant autant de catégories vides que nécessaire, les λ -expressions et les règles de composition n'étant modifiées que marginalement, soit au contraire en chargeant les règles de composition, grâce à des combinateurs *ad hoc*, de produire la représentation voulue, à partir d'un arbre syntaxique le plus simple possible.

Dans tous les cas, on considèrera (1) que les problèmes morphologiques sont résolus (*voir* et *voit* sont le même terminal), et (2) que le fait qu'il s'agisse d'un cas de verbe à contrôle sujet est déterminé par la syntaxe.