

Correction du Partiel THL

THÉORIE DES LANGAGES

EPITA – Promo 2012 – Sans documents ni machine

Janvier 2010 (1h30)

Correction: Le sujet et sa correction ont été écrits par Akim Demaille et Jonathan Fabrizio. Le best-of est tiré des copies des étudiants. Les erreurs, en particulier d'orthographe, sont les leurs !

Barème: Toutes les questions sont notées sur 4, et c'est ainsi qu'il faut lire les barèmes qui suivent. Se reporter à la feuille de calcul pour voir les coefficients donnés aux questions.

Bien lire les questions, chaque mot est important. Écrire court, juste, et bien. Une argumentation informelle mais convaincante est souvent suffisante.

Les questions à choix multiples, numérotées Q.1, Q.2 etc., sont à répondre sur les formulaires de QCM ; aucune réponse manuscrite ne sera corrigée. Il y a exactement une et une seule réponse juste pour ces questions. Si plusieurs réponses sont valides, sélectionner la plus restrictive. Par exemple s'il est demandé si 0 est *nul*, *non nul*, *positif*, ou *négatif*, sélectionner *nul* qui est plus restrictif que *positif* et *négatif*, tous deux vrais. Ne pas oublier de renseigner les champs d'identité.

1 Incontournables

Chaque erreur aux trois questions suivantes retire 1/6 de la note finale. Avoir tout faux divise donc la note par 2.

Correction: D'assez nombreux étudiants n'ont pas été capables de mettre les réponses sur le formulaire, mais l'ont fait dans leur copie. Dans ce cas, ils ont eu faux aux trois questions.

Q.1 Si le langage L est tel que $L = L_1 \cup L_2$ où L_1 et L_2 sont rationnels, alors il est rationnel.

✓ vrai

✗ faux

Correction: Oui, bien entendu, l'union fait partie des opérations rationnelles par définition.

Q.2 Si L_1, L_2 sont des langages finis, alors $\{u^n v^n \mid u \in L_1, v \in L_2, n \in \mathbb{N}\}$ est rationnel.

✗ vrai

✓ faux

Correction: Bien sûr que non, comme le montre le cas bien connu de $L_1 = \{a\}, L_2 = \{b\}$.

Q.3 Si une grammaire hors-contexte est ambiguë, alors il existe un automate fini non-déterministe qui en reconnaisse le langage.

✗ vrai

✓ faux

Correction: Faux. Aucun rapport entre les deux parties de cette question. Une grammaire hors-contexte n'engendre pas nécessairement un langage rationnel. Un automate reconnaît un langage rationnel, indépendamment de son déterminisme.

2 Contrôle

Q.4 L'expression rationnelle étendue $[-+]^2[0-9A-F]^+([-+/*] [-+]^2[0-9A-F]^+)^* n'$ engendre pas :

- $-DEAD + BEEF$ $+ - AA - + BB * + CC - * DD$
- $++ AB +++ DE$ $AA + BB - CC * DD$

Correction: La question était rédigée pour $[-+]^* [0-9A-F]^+ ([-+/*] [-+]^* [0-9A-F]^+)^*$, i.e., permettait autant de signes optionnels que voulu, et non au plus un. Si bien que les réponses (b) et (c) étaient toutes deux fausses. Par conséquent, cette question ne compte pas.

Q.5 Qui est l'inventeur de SLR(1) et LALR(1) ?

- Frank DeRemer Noam Chomsky Donald Knuth Stephen Kleene

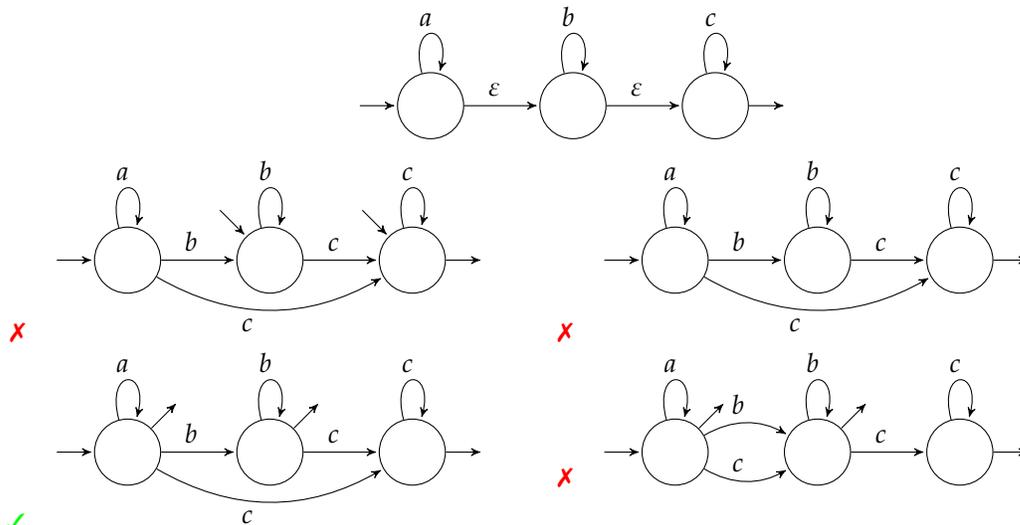
Correction: Frank DeRemer.

Q.6 Combien existe-t-il de nombres hexadécimaux de quatre chiffres ?

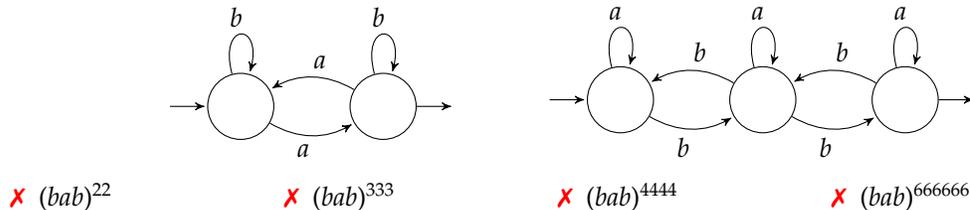
- 4^2 2^4 2^{16} 4^{16}

Correction: $16^4 = (2^4)^4 = 2^{(4 \times 4)} = 2^{16}$.

Q.7 Quel est le résultat d'une ϵ -clôture *arrière* sur l'automate suivant ?

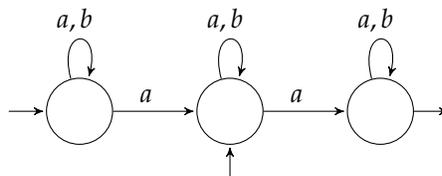


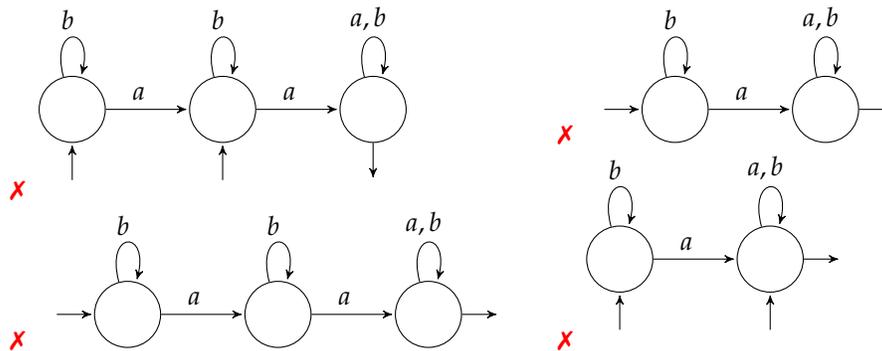
Q.8 Quel mot est reconnu par l'automate produit des deux automates suivants ?



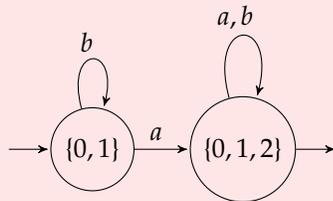
Correction: L'automate produit calcule l'intersection des langages. Le premier automate veut un nombre impair de a , le second un nombre pair non nul de b . Donc, évidemment, (b) .

Q.9 Déterminer l'automate suivant.





Correction: Ne pas oublier qu'il faut commencer par un état initial qui représente tous les états initiaux du nondéterministe.



Q.10 Quelle technique est bien adaptée à l'écriture de parseurs à la main ?

- LL(1)
- LR(0)
- SLR(1)
- LALR(1)

Correction: (a).

Q.11 Pour tout langage rationnel, il existe une grammaire

- sensible au contexte
- linéaire à gauche
- hors-contexte
- gauche
- ambiguë à gauche

Correction: Une grammaire rationnelle est par définition une grammaire linéaire à gauche, ou linéaire à droite. Donc (b) est vérifiée. (a) et (c) également, mais elles sont plus générales que (b). (c) ne veut rien dire. (b).

Q.12 Dans une analyse classique en utilisant Yacc et Lex, yyparse et yylex sont appelées respectivement :

- une fois et une fois
- fois
- fois
- sieurs fois
- une fois et plusieurs
- plusieurs fois et une
- plusieurs fois et plu-

Correction: (b).

3 Parsage LL

Considérons G , l'extrait de grammaire suivant, où les terminaux sont écrits entre guillemets.

```

stmt: stmt_if | stmt_print
stmt_if: "if" exp "then" stmt "fi"
        | "if" exp "then" stmt "else" stmt "fi"
stmt_print: "print" "(" args ")"
args: exp "," args | exp
exp: "number" | "boolean"
    
```

1. Quel est le type de Chomsky de cette grammaire ? Justifier.

Correction: Toutes les règles sont hors-contexte, et il y a de nombreuses règles non-rationnelles (comme la troisième par exemple). Donc cette grammaire est hors-contexte.

Best-of:

- Cette grammaire pose le problème du dangling else et est bien connue pour être une grammaire ambiguë
- C'est une grammaire infinie, linéaire à gauche. Donc c'est la grammaire rationnelle.
- Cette grammaire est de type (iii), ambiguë car elle comporte des conflits shift/reduce
- La grammaire est de type 2, elle est hors-contexte ou CS (contexte sensitive)
- Cette grammaire est de type 4 car c'est une grammaire hors-contexte contenant des terminaux précis, c'est-à-dire non définis par une expression rationnelle
- La grammaire est ici sensible au contexte droit
- Pour la non-rationalité, on remarque que `args` peut s'écrire sous la forme `args : exp(",exp)*` ce qui n'est pas rationnel !

Q.13 Combien de règles comporte ce bout de grammaire ?

X 5

X 6

X 9

X 14

Correction: Ne pas oublier que `|` est une abréviation pour introduire une règle qui a la même partie gauche que la règle précédente : 9 règles.

2. Quel est le type de Chomsky du langage engendré par cette grammaire ? Justifier.

Correction: La présence des parenthèses dans `stmt_print` ne pose pas de problèmes particuliers, puisqu'elles ne s'imbriquent pas. En revanche, on a bien un enchaînement des paires if/fi qui interdit à ce langage d'être rationnel. La grammaire étant hors-contexte, ce langage est hors-contexte.

Best-of:

- Le langage engendré est de type LL(5) car dans la règle `stmt_if` l'indétermination demeure jusqu'au 5^{em} token
- La grammaire est hors-contexte, donc régulière
- Le langage engendré est de type 5 car il est fini
- Le langage engendré est non déterministe

3. Cette grammaire n'est pas LL(1). Quelles règles posent problème ? Expliquez pour chacune le problème (calculer au besoin les ensembles NULLABLE, FIRST et FOLLOW des règles).

Best-of:

- La règle `stmt_if` pose problème car on rencontre une ambiguïté lorsque l'on doit choisir entre le "fi" et le "else"
- ... conflits shift/reduce ...

4. En corrigeant les problèmes identifiés à la question 3, proposer une grammaire G' équivalente (i.e., sans changer le langage engendré) qui soit LL(1).

Best-of:

- `args : args ", " exp | exp`
- `args : exp | exp ", " args`

5. Un utilisateur de ce langage se plaint du `fi`, dont il trouve l'usage trop contraignant. Est-il vraiment nécessaire pour votre parseur LL(1) ? Justifier.

Best-of:

- le `fi` est nécessaire pour le parseur LL(1) car il permet de soulever des ambiguïtés
- OUI ! Le "fi" est très important sinon le problème connu du if-else dont je ne sais plus le nom exact mais déjà tombe en question à un précédent partiel
- Cependant même si le parseur devrait pouvoir se retrouver avec cette grammaire sans "fi", le programmeur lui n'y arrivera plus quand il commencera à faire de long programme
- Non, on peut mettre des autres symboles

6. Donner en pseudo-code la routine de l'analyseur prédictif récursif descendant LL(1) de G' pour le non-terminal `stmt-if`.

4 À propos de ce cours

Nous nous engageons à ne pas tenir compte des renseignements ci-dessous pour noter votre copie. Ils ne sont pas anonymes, car nous sommes curieux de confronter vos réponses à votre note. En échange, quelques points seront attribués pour avoir répondu. Merci d'avance.

Répondez sur les formulaires de QCM qui vous sont remis. Vous pouvez cocher plusieurs réponses par question.

Q.14 Prises de notes

- a. Aucune
- b. Sur papier
- c. Sur ordinateur à clavier
- d. Sur ardoise
- e. Sur le journal du jour

Q.15 Travail personnel

- a. Rien
- b. Bachotage récent
- c. Relu les notes entre chaque cours
- d. Fait les annales
- e. Lu d'autres sources

Q.16 Ce cours

- a. Est incompréhensible et j'ai rapidement abandonné
- b. Est difficile à suivre mais j'essaie
- c. Est facile à suivre une fois qu'on a compris le truc
- d. Est trop élémentaire

Q.17 Ce cours

- a. Ne m'a donné aucune satisfaction
- b. N'a aucun intérêt dans ma formation
- c. Est une agréable curiosité
- d. Est nécessaire mais pas intéressant
- e. Je le recommande

Q.18 L'enseignant

- a. N'est pas pédagogue
- b. Parle à des étudiants qui sont au dessus de mon niveau
- c. Me parle
- d. Se répète vraiment trop
- e. Se contente de trop simple et devrait pousser le niveau vers le haut