

Question 1 Expliquer brièvement ce qu'est la loi de Zipf et quelles en sont les conséquences pour la modélisation du langage.

Question 2 Qu'est-ce qu'un morphème ?

- Donner une définition.
- Illustrez par quelques exemples de votre choix.
- Quelles techniques algorithmiques utilise-t-on pour réaliser l'analyse morphologique automatiquement ?
- Donnez quelques cas où ces techniques sont utiles en traitement automatique des langues.

Question 3

- Soit la grammaire probabiliste suivante :

$S \rightarrow GN\ GV\ (1.0)$

$GN \rightarrow D\ N\ (0.6)$

$GN \rightarrow D\ N\ GP\ (0.4)$

$GV \rightarrow V\ GN\ (0.2)$

$GV \rightarrow V\ GN\ GP\ (0.8)$

$GP \rightarrow P\ GN\ (1.0)$

$D \rightarrow le\ (0.5) \mid l'\ (0.5)$

$V \rightarrow voit\ (1.0)$

$N \rightarrow astronome\ (0.5) \mid télescope\ (0.3) \mid chien\ (0.2)$

$P \rightarrow avec\ (1.0)$

- Étant donnée cette grammaire, quelles sont les analyses possibles pour la phrase *L' astronome voit le chien avec le télescope* ?
- Donnez le calcul qui vous permet de déterminer l'analyse la plus probable donnée par cette grammaire à cette phrase.
- ✗ • Le résultat vous paraît-il satisfaisant ? pour élaborer votre réponse, comparez avec le résultat le plus plausible pour la phrase *Le chien voit l'astronome avec le télescope* ?

- ✕ • Donner un algorithme de transformation d'une grammaire d'arité quelconque en grammaire binaire. Et illustrez son déroulement sur la grammaire donnée en exemple. Dans quel cas et pourquoi utilise-t-on un algorithme de ce type ?
- Donner le pseudo-code d'un algorithme d'analyse qui permet de déterminer la meilleure analyse syntaxique possible pour une phrase donnée. Illustrez le déroulement de cet algorithme sur l'exemple donné ci-dessus. (On demande de donner le chart de l'analyse) ainsi qu'un commentaire sur la manière dont vous l'obtenez.

Question 4

- Dans un casino malhonnête, le jeu de roulette est de temps en temps truqué. Le jeu de roulette comporte 37 numéros. La plupart du temps la chance de tomber sur 0 est de $\frac{1}{37}$ comme pour tous les autres numéros, lorsque le jeu est truqué, on a une chance sur 10 de tomber sur zéro, alors qu'on a 1 chance sur 40 de tomber sur un autre nombre. Le casino passe du jeu équitable au jeu truqué avec une probabilité de 0.2 et repasse au jeu équitable avec une probabilité de 0.1. De plus, le casino a une chance sur deux de démarrer son jeu dans l'état truqué. Le changement du jeu correspond donc à un processus markovien et les résultats des tirages change selon que le jeu est truqué ou pas.
 - Donner une représentation graphique du HMM qui correspond à ce jeu de roulette truqué
 - Quels sont les états cachés ? quels sont les symboles observés ?