

2022-2023 Master 1 informatique HAI815I langage naturel 1 (syntaxe)
Examen session 1 : mardi 23 mai 2023 09 :00 FDS Salle 36.3
Durée 2 heures. Sujet : 1 page.
Aucun document autorisé. Calculatrice autorisée.

I — MODÈLES DE MARKOV CACHÉS

On souhaite définir un Modèle de Markov caché (HMM) avec les états "Initial, déterminant, nom, adjectif, verbe, pronom, Final" pour étiqueter des phrases construites sur le vocabulaire figurant dans les phrases suivantes, l'étiquetage étant déterminé par le chemin de Viterbi. [Il est assez ridicule de faire des statistiques et des probabilités sur si peu d'exemples, mais cet exercice a pour but de vérifier si vous avez compris tout en évitant de longs calculs.]

- (a). Ecrire le HMM avec les probabilités induites par les exemples suivants, en expliquant au préalable comment sont calculées les probabilités (apprentissage par (ré)estimation des probabilités) :
- (1) La bouchère sale une tranche.
det nom vb det nom
 - (2) La bouchère sale la tranche.
det nom adj pro vb
 - (3) Une bouchère sale une tranche.
det nom vb det nom
- (b). Quel étiquetage votre HMM attribue-t-il à la phrase suivante ? On justifiera sa réponse et on précisera les calculs. Ceux qui n'utiliseront pas l'algorithme de Viterbi auront au plus la moitié des points attribués à cette question.
- (4) Une bouchère sale la tranche.

II — GRAMMAIRES FORMELLES POUR LE LANGAGE NATUREL

- (a). Donnez une grammaire hors contexte qui engendre toutes les phrases écrites avec les mots des phrases de l'exercice précédent grammaticalement correctes (y compris du point de vue de l'accord en genre et en nombre à l'intérieur du groupe nominal et de l'accord verbe sujet en nombre). On fera attention à ce que "la tranche" soit un groupe nominal mais aussi un groupe verbal (pronom objet "la" suivi du verbe transitif "tranche"), à ce que chaque mot puisse apparaître avec les diverses catégories, comme dans l'exercice précédent.
- (b). Existe-t-il une grammaire régulière qui engendre le même langage ? Expliquez pourquoi.
- (c). Analyser la phrase "(4) : Une bouchère sale la tranche." par l'algorithme de Cocke-Kasami-Younger. On commencera par expliquer cet algorithme. Ceux qui utiliseront une autre méthode auront au plus la moitié des points attribués à cette question. Quelles analyses syntaxiques trouvez-vous ? Expliquez-le lien avec la dernière question de l'exercice précédent.
- (d). Ecrivez votre grammaire hors-contexte comme une DCG (Definite Clause Grammar) en Prolog (avec différences de listes, comme fait en TP).
- (e). Les verbes "saler" et "trancher" ont obligatoirement un sujet animé et un complément d'objet inanimé. Modifiez votre DCG afin que cette restriction soit prise en compte et que "la tranche sale(vt) la bouchère", "la tranche la sale(vt)" ne soient plus reconnues par votre grammaire. Vous procéderez comme vous l'aviez fait en TP pour imposer l'accord en genre et en nombre.

III — SYSTÈME D'INFÉRENCES (MINI-PROJET INFÉRENCES ET EXPLICATIONS) / MAX 2 PAGES

- (a). Quels sont les différents types d'inférence que l'on peut réaliser automatiquement ? Pour chacun d'entre eux, indiquez le niveau de risque de se tromper.
- (b). Comment évaluer le niveau de pertinence d'une inférence donnée ?
- (c). Si on connaît la réponse à une question, les inférences peuvent servir d'explication. Expliquer comment connaître la réponse peut aider à orienter le calcul d'explications