[2 points] Soit les deux descriptions de cours suivantes, une du programme d'informatique et l'autre du programme d'études littéraires et culturelles, à l'Université de Sherbrooke :

Informatique	Études littéraires et culturelles
Formaliser les structures de données, comparer et choisir les meilleures mises en oeuvre des structures en fonction du problème à traiter.	Étudier, d'un point de vue épistémologique, les notions de littérature et de culture. Étudier les rapports qu'elles entretiennent entre elles.

On a donc deux corpus, un pour la catégorie informatique, l'autre pour les études littéraires et culturelles. Soit la nouvelle description de cours suivante :

> Étudier les notions de base en théorie des graphes. Étudier les structures de données externes.

Supposez l'utilisation du vocabulaire suivant : {"structures", "données", "Étudier", "culture", "'notions", "."}.

(a) Quelle est la distribution unigramme associée à la catégorie informatique ? Utiliser une constante de lissage $\delta = 0.1$.

(b) Quelle est la distribution unigramme associée à la catégorie études littéraires et culturelles ? Utiliser aussi une constante de lissage $\delta = 0.1$.

(c) À l'aide des distributions unigrammes calculées en (a) et (b), et en supposant une distribution a priori uniforme sur les catégories (c.-à-d. $P(C = \text{informatique}) = P(C = \text{études littéraires et culturelles}) = 0.5$.

Déterminez dans quelle catégorie la nouvelle description serait classifiée, par un classifieur bayésien naïf multinomial.