

TD 2: TRANSDUCTEURS PONDÉRÉS À ÉTATS FINIS

Exercice 1.

Représenter graphiquement les transducteurs correspondant aux situations suivantes. Pour chaque transducteur définir les alphabets d'entrée et de sortie.

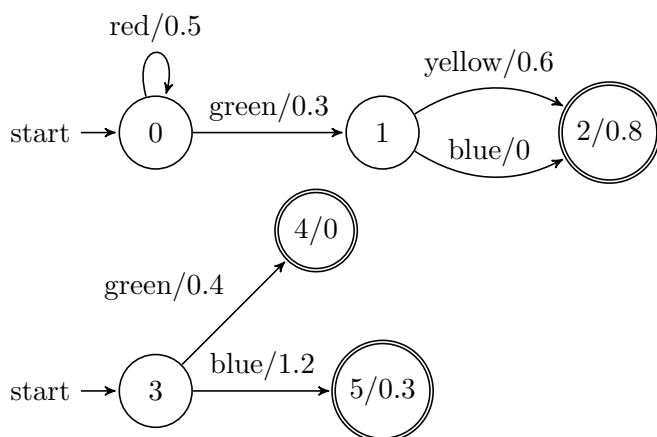
1. Une machine qui rit: elle prend en entrée la chaîne de caractère **rire** et le traduit en rire, c'est-à-dire l'ensemble suivant: $\{ha^n | n \geq 1\}$
2. Une machine qui décrit certains adjectifs: si ils sont masculins ou féminins, singuliers ou pluriels. On pourra prendre comme exemple *petit, belle, froid, ennuyeux*.
3. Une machine qui supprime les balises dans un fichier texte "<>"

Exercice 2.

1. Est-ce que les relations $\mathcal{R}_1 = (a, b)^*$ et $\mathcal{R}_2 = (\epsilon, c)^*$ sont régulières ?
2. Est-ce que la relation $\mathcal{S} = (a, bc)^*$ est régulière ?
3. Est-ce que la relation $\mathcal{T} = \mathcal{R}_1 \circ \mathcal{R}_2$ est régulière ? La définir.
4. Est-ce que la relation $\mathcal{U} = \mathcal{R}_1 \cap \mathcal{R}_2$ est régulière ?

Exercice 3.

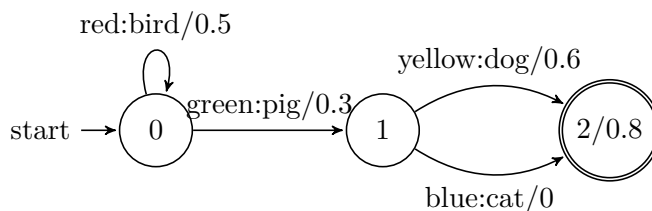
Soit les automates pondérés suivants:



1. Sur quels semi-anneaux les pondérations des automates ci-dessus peuvent-elles être définies ?
2. Représenter graphiquement (ou donner la fonction de transition de) l'union des deux automates.

Exercice 4.

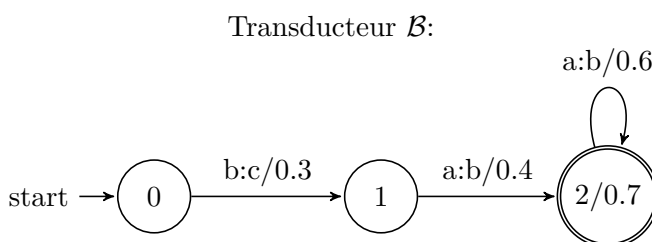
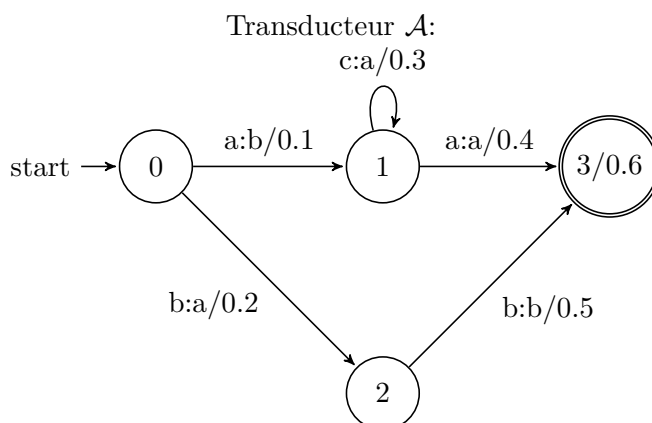
Soit le transducteur pondéré suivant:



1. Représenter graphiquement (ou donner la fonction de transition de) le transducteur transposé.
2. Représenter graphiquement (ou donner la fonction de transition de) la projection sur le domaine.

Exercice 5.

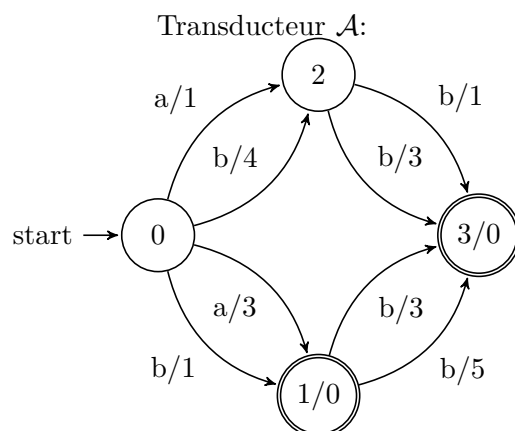
Soit les transducteurs pondérés suivants:



1. Donner les alphabets d'entrée et de sortie de chaque transducteur.
2. On considère que les pondérations des transducteurs sont définies sur le semi-anneau tropical, donner la composition de $\mathcal{A} \circ \mathcal{B}$?

Exercice 6.

Soit le transducteur pondéré, dont les pondérations sont définies sur le semi-anneau tropical:



1. Déterminer les états jumeaux, que concluez-vous ?
2. Déterminiser \mathcal{A} .

Exercice 7.

On cherche à créer un modèle de langue. Pour cela, on récupère un grand nombre de documents écrits (livres, sites web, twitter, etc...) et on recense les séquences de n mots consécutifs. Ces séquences sont appelées n -grams. Un exemple de 3-grams est donné ci dessous pour l'anglais.

frequency	word1	word2	word3
1419	much	the	same
461	much	more	likely
432	much	better	than
266	much	more	difficult
235	much	of	the
226	much	more	than

1. Pour chaque 3-grams, donner la log-vraisemblance associée. On considère, en grossière approximation, qu'il n'y a pas d'autre 3-grams.
2. Proposer un automate reconnaissant ces 3-grams, les pondérations étant définies sur le semi-anneau $(\mathbb{R} \cup \{-\infty; +\infty\}, \oplus_{\log}, +, +\infty, 0)$