GENERATION DE MOTS ET DE MUSIQUE

Année 2025 - 2026

Ambroise CIRIONI, Mandélya LANGLADE, Clara FERNANDO : élèves de classe de terminale.

Établissement : Lycée Olympe de Gouges de Montech.

Enseignant.e(s) : Adeline Cavailles, Chloé Samson et Fabien Bourg**.**

Chercheur.Chercheuse(s) : Lucas Monteiro et Caroline , Université Paul-Sabatier ,Toulouse.

On cherche à savoir combien de mots il est possible de former avec le mot **ABRACADABRA** en sachant que **A** ne peut être suivi de **B;C;D**, **B** ne sera suivi que de **R** et **C;D;R** ne peuvent être suivis que de **A**.

Mot de 2 lettres

↱	**A**	**B**	**R**	**C**	**D**
A	**0**	**1**	**0**	**1**	**1**
B	**0**	**0**	**1**	**0**	**0**
R	**1**	**0**	**0**	**0**	**0**
C	**1**	**0**	**0**	**0**	**0**
D	**1**	**0**	**0**	**0**	**0**

Mot de base

ABRACADABRA

On crée une matrice qui cherche à savoir s'il est possible d'avoir une lettre après une autre.

Si oui, on met 1, si non alors on met 0.

Cette matrice se lit de la sorte: le 1 dans la première ligne, deuxième colonne signifie qu'il y a un seul mot de 2 lettres qui commence par A et qui finit par B.

Soit : Ab. Dans la même idée, on a: Ac, Ad, Br, Ra, Ca, Da

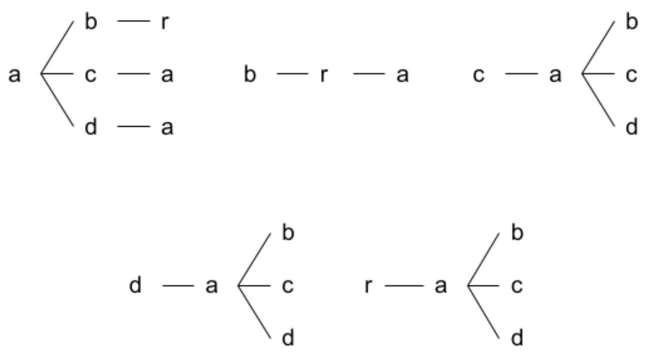
On la multiplie par elle-même pour avoir une lettre en plus dans la longueur du mot.

↱	A	B	R	C	D
A	2	0	1	0	0
B	1	0	0	0	0
R	0	1	0	1	1
C	0	1	0	1	1
D	0	1	0	1	1

La matrice précédente comptait le nombre de mots possibles mais aussi s'il était possible d'avoir une lettre après une autre. Celle-ci ne compte que le <u>nombre possible de mots</u> à former en gardant les conditions initiales des suites de lettres.
On additionne toutes les valeurs pour connaître le nombre maximal de mots de longueur 3.
2+1*11=13. Soit 13 mots possibles et différents

ABA, ACA, ABR, BRA, CAB, CAC, CAD, DAB, DAC, DAD, RAB, RAC, RAD

<u>Avec un arbre de dénombrement</u>

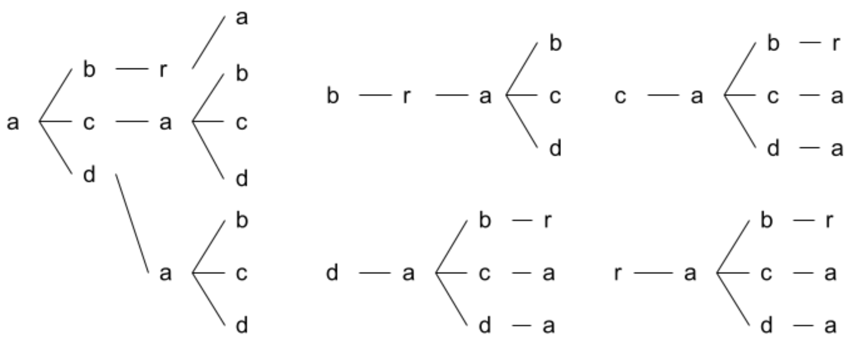


On vérifie avec une longueur de 4 lettres, on multiplie notre matrice par elle-même 3 fois. Si on additionne les valeurs de la matrice et qu'on trouve le même nombre que les issues de l'arbre de probabilité, comme précédemment, alors on pourra conclure.(1)

↱	A	B	R	C	D
A	1	2	0	2	2
B	0	1	0	1	1
R	2	0	1	0	0
C	2	0	1	0	0
D	2	0	1	0	0

Pour un mot de longueur de 4 lettres, il y avait : 1*7+2*6= 19 mots possibles et différents.

Vérifions désormais avec l'arbre de dénombrement:



19 mots de 4 lettres

Soit, il faut multiplier la matrice par elle-même la longueur qu'on souhaite -1 pour avoir le nombre de mots possibles(2). Dans notre cas, **ABRACADABRA** a 11 lettres, donc on multiplie la matrice par elle-même 10 fois.

↱	A	B	R	C	D
A	**56**	**33**	**22**	**33**	**33**
B	**22**	**12**	**9**	**12**	**12**
R	**33**	**22**	**12**	**22**	**22**
C	**33**	**22**	**12**	**22**	**22**
D	**33**	**22**	**12**	**22**	**22**

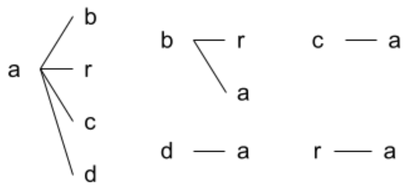
9+12*6+22*11+33*6+56=577

On peut donc former 577 mots différents avec **ABRACADABRA**

Si on change deux lettres de position **ABRACADABAR**, **A** pourra aussi être suivi d'un **R** et **B** d'un **A**, pourrons-nous former plus de mots ?
La démarche reste la même, il faudra multiplier la nouvelle matrice associée au graphe(3) par elle-même.

Mot de 2 lettres **Arbre de probabilité**

↱	A	B	R	C	D
A	0	1	1	1	1
B	0	0	1	0	0
R	1	0	0	0	0
C	1	0	0	0	0
D	1	0	0	0	0



On la multiplie par elle-même 10 fois pour avoir un mot de 11 lettres

↱	A	B	R	C	D
A	1120	257	525	257	25
B	525	316	429	316	31
R	257	268	316	268	26
C	257	268	316	268	26
D	257	268	316	268	26

(4)(

25+26*3+31+257*5+268*6+316*5+429+525*2+1120
= 7 206

On peut donc former 7 206 mots différents de longueur 11 avec **ABRACADABAR**

Notre hypothèse est bien confirmée(5).

Il y a plus de possibilité d'avoir une lettre après une autre dans ABRACADABAR, donc on peut former plus de mots.
On peut donc faire le lien avec l'IA, il y a une différence majeurs entre IA et notre problème, car nous avons 2 méthode différente, notre méthode consiste à faire des prédiction à l'aide de probabilité fix déjà défini et l'IA utilise des réseau de neurone qui permet de faire des similarité entre les mots et donc de deviner de façon plus intéressante la suite d'une phrase.

Nous pouvons donc générer grâce à un programme python des phrase de texte grâce aux probabilités dans les matrices. Comme nous voulons des probabilité et pas des matrices d'adjacence, nous prenons la même matrice que la 1 (ABRACADABRA) et nous lui appliquons les probabilités avant de la multiplier par elle même(6).

↱	A	B	R	C	D
A	0	0.5	0	0.25	0.25
B	0	0	1	0	0
R	1	0	0	0	0
C	1	0	0	0	0
D	1	0	0	0	0

On a donc une matrice de probabilité, elle permet de connaître la probabilité de commencer et de finir par une lettre pour la puissance n-1 qui est aussi le nombre n qui correspond au nombre de caractères.

Pour générer, on va tirer aléatoirement un objet de la matrice puis à l'aide de la matrice de probabilité à la puissance n, on va déterminer les probabilités des objets suivants puis en tirer un parmi tous.

On peut donc avoir notre phrase avec n objets et cela permet de faire une phrase logique bien que sans sens propre.(7)

Pour générer la musique, on convertit la musique en note puis on génère une phrase puis on convertit les notes obtenues en son, ça nous donne une musique.

https://github.com/enzoambroise/Math-en-Jean-2026-Generation-de-mots-et-de-music

Notes d'édition

(1) Les auteurs de l'article veulent ici vérifier, sur cet exemple de recherche du nombre de mots de 3 lettres (puis de 4) en appliquant la règle donnée par ABRACADABRA, que la matrice obtenue en multipliant la matrice initiale A deux fois (ou trois fois) de suite par elle-même (c'est-à-dire en calculant son carré, ou son cube) donne ce nombre en additionnant tous ses coefficients. Or, il est clair que l'arbre de dénombrement donne bien tous ces mots de 3 (ou 4) lettres. Ce qui permet de faire cette vérification.

(2) Si on prend autre mot que ABRACADABRA, on peut construire une matrice initiale A en utilisant les mêmes règles que pour ABRACADABRA ; si on veut connaître le nombre de mots de n lettres construits avec ces règles, on calcule A^{n-1}et on fait la somme des coefficients de A^{n-1}.

(3) On peut donc ici faire la conjecture suivante : si on prend un mot quelconque et qu'on lui associe la matrice A construite comme pour ABRACADABRA, alors, le nombre de mots de n

lettres construits avec ces règles est donné par la somme des coefficients de A^{n-1}. Cette conjecture n'est pas démontrée ici mais seulement vérifiée sur quelques exemples.

(4) Le graphe dont il est question ici est celui de la relation « x est en relation avec y si y suit x dans le mot de départ ». La matrice construite s'appelle la « matrice d'adjacence » du graphe.

(5) Le calcul de A^{10} semble faux, mais le principe est le bon.

(6) On remplace dans la matrice initiale et dans la ligne de la lettre x les 1 et les 0 par les fréquences avec lesquelles la lettre y suit la lettre x dans la matrice de départ. Ainsi, dans l'exemple ABRACADABRA, il y a 3 lettres qui suivent A : B avec la fréquence de ½, D avec la fréquence ¼ et C avec la fréquence ¼.

(7) La méthode proposée pour obtenir les mots de n lettres à partir d'un mot donné (comme ABRACADABRA) semble être la suivante :

-on choisit une lettre au hasard dans le mot donné ;

-on utilise la matrice A^{n-1}pour écrire tous les mots de n lettres : chacun de ces mots est obtenu avec une probabilité donnée par la matrice B^{n-1}où B est la matrice de probabilité obtenue à partir de la matrice A. Ceci est une conjecture qui n'est pas démontrée.