Cours

Modèles de Markov Cachés (MMC)

(En anglais: Hidden Markov Models : HMM)

Les modèles de Markov observables

- Temps discret : $t = 1, 2, \ldots, T$
- Une chaine ou *processus de Markov* est une **séquence** de variables aléatoires:

$$\{X_1, X_2, \ldots, X_T\}$$

- Ces variables aléatoires prennent leurs valeurs dans l'espace d'états :

$$S = \{q_1, q_2, \ldots, q_N\}.$$

- On dit que le processus stochastique **émet des séquences d'états** $q_1, q_2, \ldots, q_T$
- Un processus stochastique est dit ***markovien*** s'il vérifie les deux propriétés suivantes :

  1. Historique limité
  2. Indépendance du temps

Historique limité

- L'état courant dépend d'un nombre constant d'états précédents.
- Par exemple, on parle de processus du **premier ordre** lorsque:

$$P\left(X_t = q_j \middle| X_1, X_2, \dots, X_{t-1}\right) = P\left(X_t = q_j \middle| X_{t-1}\right)$$

$\boldsymbol{X_t = q_j}$: l'état observé à l'instant t est q_j

- Autrement :

« le future ne dépend que de l'état courant et non pas du passé! »

- Donc:

$$P(q_1, q_2, \dots, q_T) = P(q_1) \times P(q_2|q_1) \times \cdots \times P(q_T|q_{T-1})$$

Indépendance du temps

- Les probabilités de transition (passage d'un état à un autre) ne varie pas avec le temps ; cela est formellement exprimé comme suit (*toujours pour un processus du premier ordre*):

$$P\left(X_t = q_j \middle| X_{t-1} = q_i\right) = a_{ij}$$

Avec:

$$a_{ij} \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1$$

$$1 \leq i, j \leq N$$

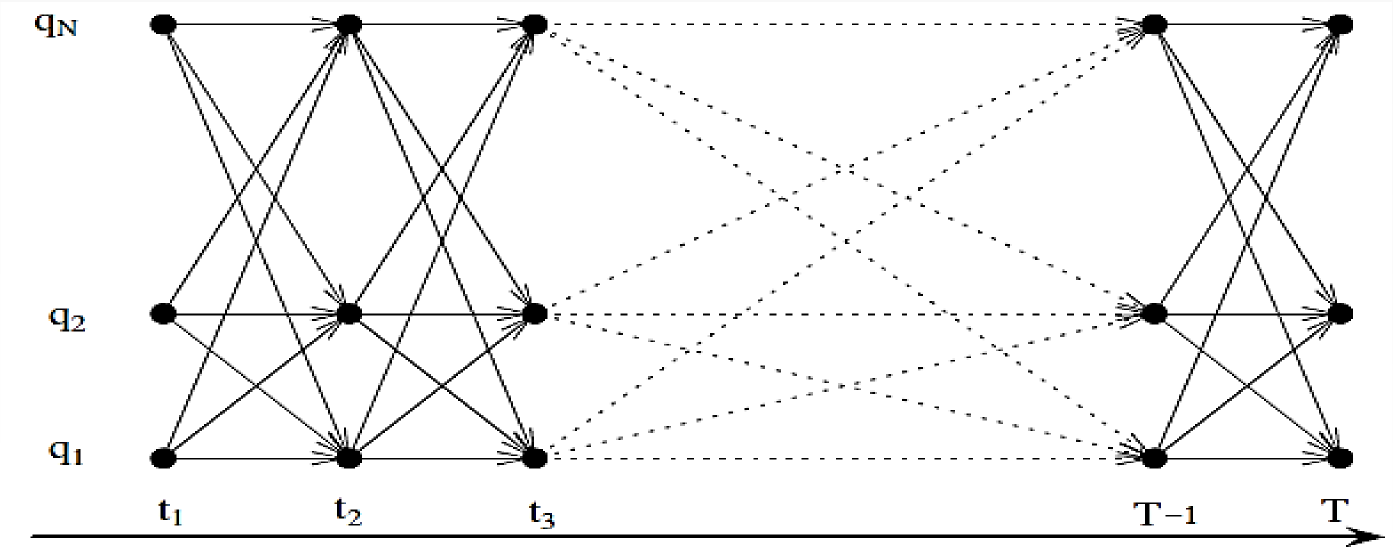
- On dénote A la matrice des probabilités de transition.
- On dit que le processus de Markov auquel nous avons affaire est **stationnaire**.

Représentation du modèle de Markov

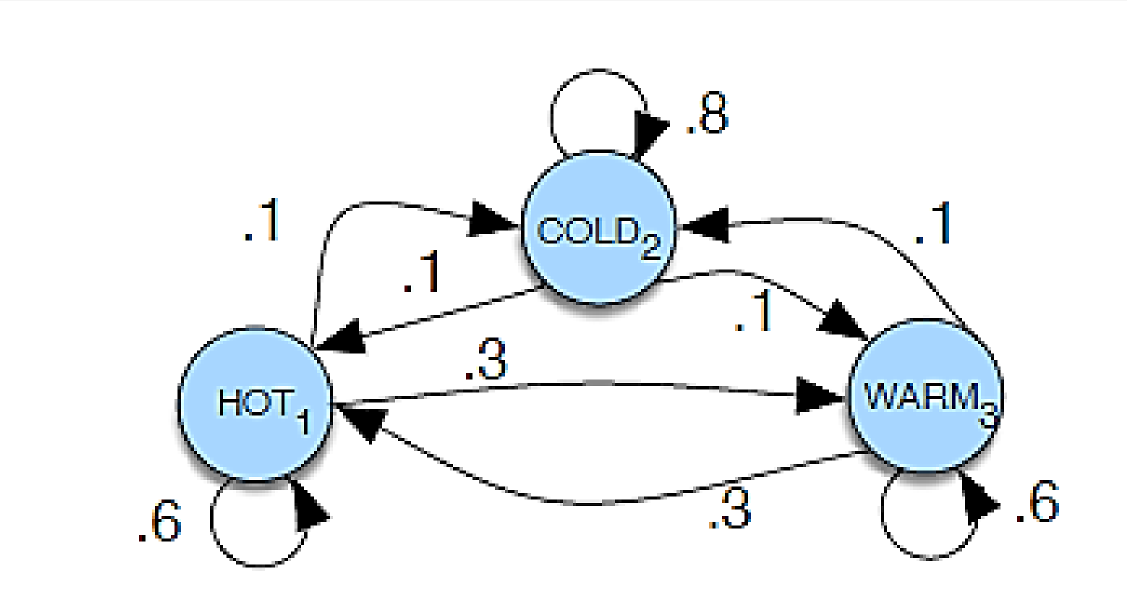
- Une chaine de Markov peut être alors vue comme un **automate probabiliste.**
- Différemment d'un automate ordinaire, l'état initial (*à l'instant t=1*) peut être n'importe quel élément dans S.
- Plus formellement, on définit la loi de probabilité de l'état initial $\Pi = \{\pi_i\}$ tel que :

$$\pi_i = P(X_1 = q_i)\ avec\ \sum_{i=1}^{N} \pi_i = 1$$

- ***L'évolution*** d'un modèle de Markov au cours du temps peut être schématisée par un treillis



Exemple d’une chaîne de Markov pour la prédiction du temps



Sachant que $\Pi = \{0.1, 0.7, 0.2\}$, trouver la probabilité des séquences:

- COLD HOT COLD HOT
- HOT HOT HOT HOT

Modèles de Markov Cachés

- La séquence d'états est **inobservable** (*d'où le nom caché du modèle*).
- on peut plutôt observer la séquence des symboles émis par les états :

$$O = O_1, O_2, \dots, O_T \,.$$

- En plus de la propriété de Markov, il y a l'**Indépendance de la sortie**:

$$P(o_i | q_1, \dots, q_T, o_1, \dots, o_T) = P(o_i | q_i)$$

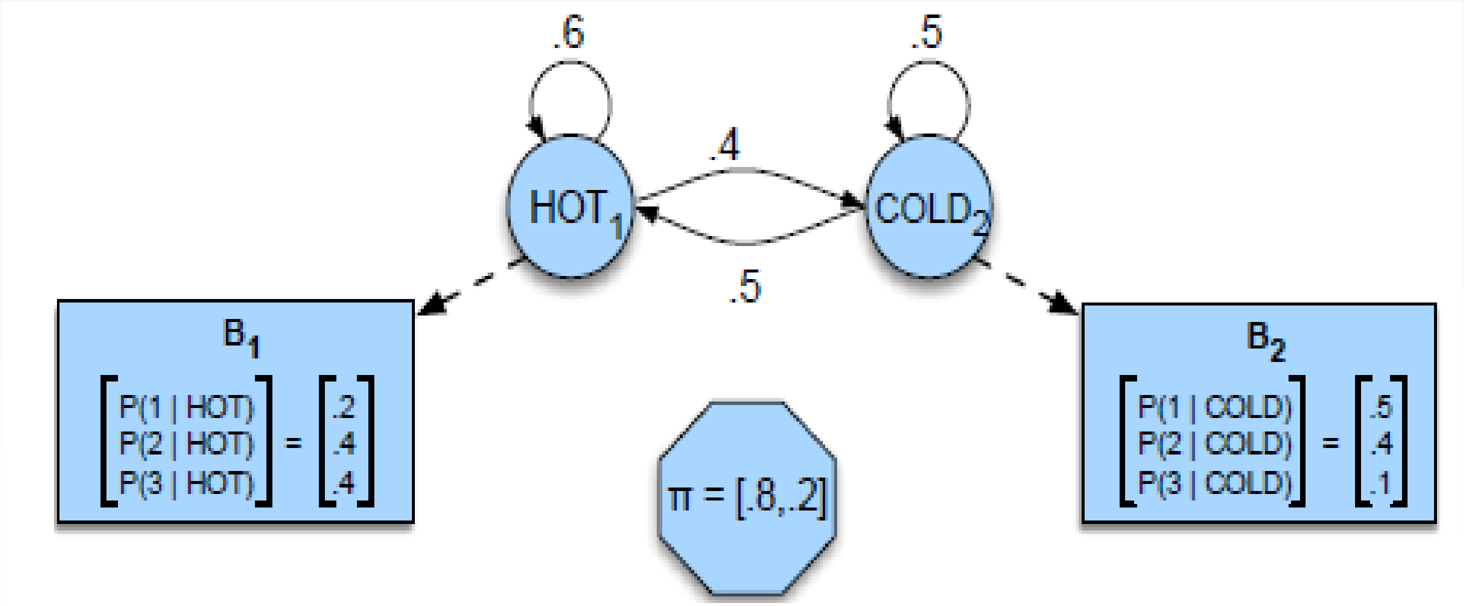
- Un modèle de Markov caché est défini par un quintuple $< S, V, \Pi, A, B >$:
  - $V = \{v_1, v_2, \dots, v_M\}$: alphabet des symboles émis par les états.
  - B dénote la matrice des probabilités d'émission ; la probabilité que l'état q_j émet le symbole v_k,$b_j(v_k)$, est définie comme suit :

    $$b_j(v_k) = P(o_t = v_k | X_t = q_j)$$

  - S,Π, A comme déjà définis .

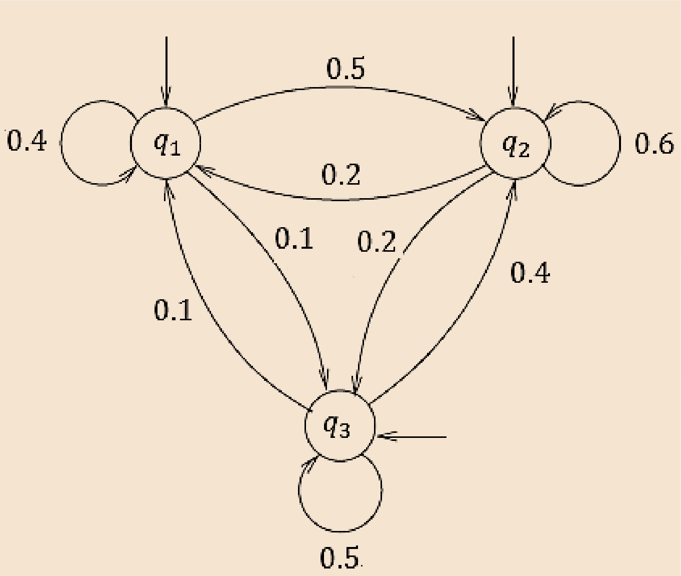
Exemple 1 (prédiction du temps)

Imaginons que vous ne pouvez pas trouver d'enregistrements de la météo à Jijel l'été 2020, mais vous trouvez l'agenda de la petite Lila qui enregistre **combien de glaces (entre 1 et 3)** elle a mangées tous les jours cet été là. Notre objectif est d'utiliser ces observations pour **estimer la température** chaque jour. Nous allons simplifier cette tâche météo en supposant il n'y a que deux sortes de jours : froid (C) et chaud (H).



Exemple 2

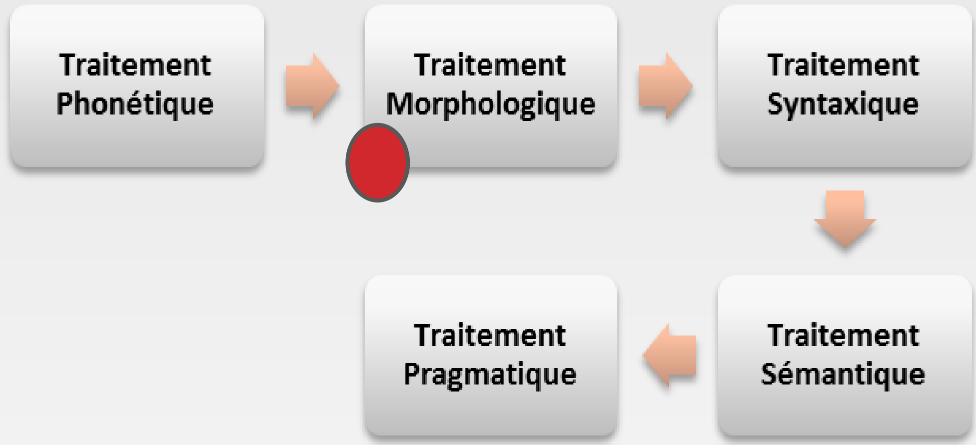
- Expérience aléatoire : lancer trois pièces de monnaie biaisées.
- S= $\{q_1, q_2, q_3\}$ tel que : $q_1 : pièce\ 1, q_2 : pièce\ 2, q_3 : pièce\ 3$
- La matrice de transition $A = \begin{bmatrix} 0.4 & 0.5 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$
- $V = \{P, F\}$ (P : Pile, F : Face)
- probabilités d'émission $B = \begin{bmatrix} 0.48 & 0.52 \\ 0.80 & 0{,}20 \\ 0.55 & 0.45 \end{bmatrix}$
- Les probabilités de l'état initial : $\pi_1 = \pi_2 = 0.3\ \pi_3 = 0.4$



Imaginons que nous avons observé **FPPFPPP,** la question qui se pose est « quelle est la suite d'états ou bien le chemin la/le plus probable ayant généré cette séquence ?».

Une application en TALN

Processus du TALN



- Génération de mots à partir de morphèmes / Analyse morphologique

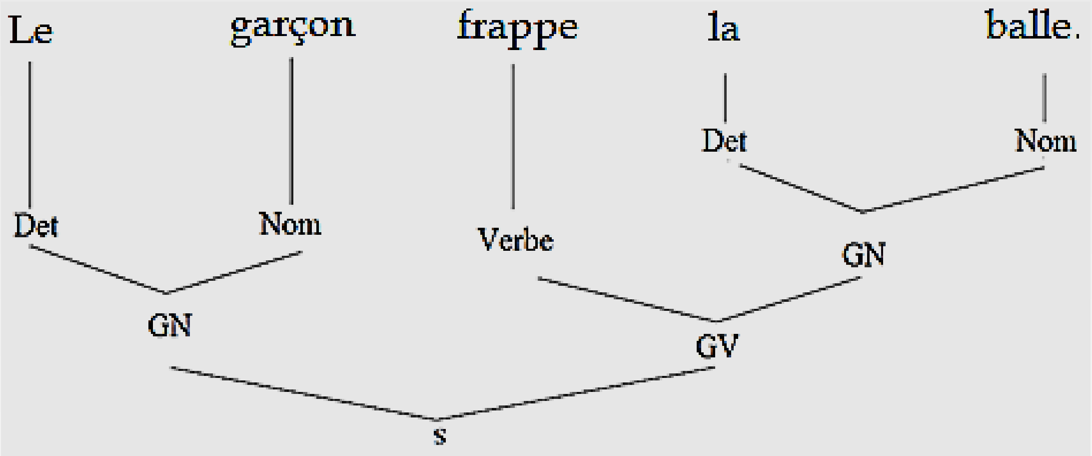
Mot	Racine	Préfixe dérivationnel	Suffixe dérivationnel	Suffixe flexionnel
intolérables	tolér	in	able	s

- **L'étiquetage morphosyntaxique** consiste à faire associer à chaque mot sa catégorie grammaticale.

Le	gros	chat	mange	la	souris	grise
article	adjectif	nom	verbe	article	nom	adjectif

L'étiquetage morphosyntaxique: intérêt ?

- Base de l'étape suivante qui est : l'analyse syntaxique.



- Aide à comprendre la sémantique d'une phrase :

Exemple **:** **"We can can the can".**

- Le mot "can" a plusieurs sens :

(1-verbe) pouvoir

(2-verbe) mettre en boîte

(3-nom) boîte.

- Après étiquetage morpho-syntaxique, on aura :

"We[pronom] can[verbe] can[verbe] the[déterminant] can[nom]".

Projection du modèle de Markov Caché sur l'application de L'étiquetage morphosyntaxique

MMC	Application en étiquetage morphosyntaxique
S	Ensemble des catégories morphosyntaxiques.
V	Ensemble des mots (le vocabulaire).
O	Une séquence de T mots observés qui prennent leurs valeurs dans V.
A	Probabilités que deux catégories morphosyntaxiques se suivent.
B	Probabilités d'observer un mot étant donnée une catégorie morphosyntaxique.
Π	La probabilité de débuter par une catégorie morphosyntaxique.

Corpus

- BD Textuelle, collection de Textes, un ensemble de documents servant d'échantillon et permettant d'extraire un ensemble d'informations utiles pour des traitements statistiques.
- Exemple:

Corpus de Brown est un **corpus annoté** avec 87 **étiquettes morphosyntaxiques**. C'est une collection de 1 million de mots provenant de **500 textes** de différents genres (journaux, romans, académiques, etc.)

Algorithme de Viterbi

- Etant donnés une observation et un modèle de Markov caché en entrée, l'algorithme de Viterbi permet de calculer la séquence d'états *optimale* qui a généré cette observation.
- Cette problématique est souvent connue par ***la problématique de décodage.***
- Notations
  - ✓ $\lambda = (A, B, \Pi)$.
  - ✓ $\delta_t(j)$ la probabilité maximale d'une observation $o_1, o_2, \dots, o_t$ sous la condition que l'état courant soit q_j à l'instant t :

$$\delta_t(j) = \max_{s_1, s_2, \dots, s_{t-1}} P(s_1, s_2, \dots, s_{t-1}, o\,1, o\,2, \dots, o\,t, \boldsymbol{s_t = q_j} | \lambda)$$

  - ✓ $\psi_t(j)$ est le chemin optimal correspondant.

1. Initialisation pour t=1

   Pour i=1 à N faire

   $\delta_1(i) = \pi_i \times b_i(o_1)$

   $\psi_1(i) = null$

   Fin Pour

2. Boucle d'induction

   t =2

   Tant que $t \leq T$ faire

   j= 1

   Tant que $j \leq N$ faire

   $\delta_t(j) = b_j(o_t) \times \max_{1 \leq i \leq N} \delta_{t-1}(i).a_{ij}$

   $\psi_t(j) = \arg\max_{1 \leq i \leq N} \delta_{t-1}(i).a_{ij}$

   $j = j + 1$

   $Fin\ TQ$

   t=t+1

   $Fin\ TQ$

3. Terminaison

Pour i=1 à N faire

$$\delta_1(i) = \pi_i \times b_i(o_1)$$

$$\psi_1(i) = null$$

	δ_1(i)	δ_2(i)	δ_3(i)	δ_4(i)	δ_5(i)
q_4:NN	0				
q_3: TO	0				
q_2: VB	0				
q_1: PPSS	0				
q_0: Start	1				
	o1:« s »	o2: I	o3:ωant	o4:to	o5:race

$\boldsymbol{\psi_1(\text{start}) = \text{Null}}$

$\boldsymbol{\psi_1(PPSS) = \text{Null}}$

$\boldsymbol{\psi_1(VB) = \text{Null}}$

$\boldsymbol{\psi_1(TO) = \text{Null}}$

$\boldsymbol{\psi_1(NN) = \text{Null}}$

j= 1

Tant que $j \leq N$ faire

$$\delta_t(j) = \mathrm{b}_j(o_t) \times \max_{1 \leq i \leq N} \delta_{t-1}(i).\, a_{ij}$$

$$\psi_t(j) = \arg\max_{1 \leq i \leq N} \delta_{t-1}(i).\, a_{ij}$$

$$j = j + 1$$

	δ_1**(i)**	δ_2**(i)**
q_4:NN	0	0×
q_3: TO	0	0×
q_2: VB	0	0×
q_1: PPSS	0	0,37 × 1 × 0,067
q_0: Start	**1**	
	o1:« s »	o2: I

j= 1

Tant que $j \leq N$ faire

$$\delta_t(j) = \mathrm{b}_j(o_t) \times \max_{1 \leq i \leq N} \delta_{t-1}(i).\, a_{ij}$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} \delta_{t-1}(i).\, a_{ij}$$

$$j = j + 1$$

	δ_1(i)	δ_2(i)	δ_3(i)
q_4:NN	0	0	0,000054×0,025×0,0012
q_3: TO	0	0	0×
q_2: VB	0	0	0,0093×0,025× 0,23
q_1: PPSS	0	**0,025**	0×
q_0: Start	**1**		
	o1:« s »	o2: I	o3:ωant

$\boldsymbol{\psi_2(PPSS)}$ **= « S »**
$\boldsymbol{\psi_2(VB)}$ **= « S »**
$\boldsymbol{\psi_2(TO)}$ **= « S »**
$\boldsymbol{\psi_2(NN)}$ **= « S »**

	δ_1**(i)**	δ_2**(i)**	δ_3**(i)**	δ_4**(i)**
q_4:NN	0	0	0.162×10^{-8}	$0 \times$
q_3: TO	0	0	0	$0.99 \times 0{,}53 \times 10^{-4} \times 0{,}035$
q_2: VB	0	0	$\mathbf{0{,}53 \times 10^{-4}}$	$0 \times$
q_1: PPSS	0	**0,025**	0	$0 \times$
q_0: Start	**1**			
	o1:« s »	o2: I	o3:ωant	o4:to

$\boldsymbol{\psi_3(PPSS) = \mathbf{PPSS}}$
$\boldsymbol{\psi_3(VB) = \mathbf{PPSS}}$
$\boldsymbol{\psi_3(TO) = \mathbf{PPSS}}$
$\boldsymbol{\psi_3(NN) = \mathbf{PPSS}}$

	δ_1**(i)**	δ_2**(i)**	δ_3**(i)**	δ_4**(i)**
q_4:NN	0	0	0.162×10^{-8}	0
q_3: TO	0	0	0	$\mathbf{0.18 \times 10^{-5}}$
q_2: VB	0	0	**0,53**$\times$ $\mathbf{10^{-4}}$	0
q_1: PPSS	0	**0,025**	0	0
q_0: Start	**1**			
	o1:« s »	o2: I	o3:$want$	o4:to

$\boldsymbol{\psi_4(PPSS)} = \mathbf{VB}$
$\boldsymbol{\psi_4(VB)} = \mathbf{VB}$
$\boldsymbol{\psi_4(TO)} = \mathbf{VB}$
$\boldsymbol{\psi_4(NN)} = \mathbf{VB}$

	δ_1(i)	δ_2(i)	δ_3(i)	δ_4(i)	δ_5(i)
q_4:NN	0	0	0.162×10^{-8}	0	$0{,}00057\times 0.18 \times 10^{-5} \times 0{,}00047 = 0{,}48\times 10^{-13}$
q_3: TO	0	0	0	$\mathbf{0.18 \times 10^{-5}}$	$0 \times$
q_2: VB	0	0	$\mathbf{0{,}53\times 10^{-4}}$	0	$0{,}00012\times 0.18 \times 10^{-5} \times 0{,}83 = \mathbf{0{,}179\times 10^{-9}}$
q_1: PPSS	0	**0,025**	0	0	$0\times$
q_0: Start	**1**				
	o1:« s »	o2: I	o3:$want$	o4:to	O5: race

$$\psi_5(PPSS) = \textbf{TO}$$
$$\psi_5(VB) = \textbf{TO}$$
$$\psi_5(TO) = \textbf{TO}$$
$$\psi_5(NN) = \textbf{TO}$$

	δ_1(i)	δ_2(i)	δ_3(i)	δ_4(i)	δ_5(i)
q_4:NN	0	0	0.162×10^{-8}	0	$0{,}48\times 10^{-13}$
q_3: TO	0	0	0	$\mathbf{0.18 \times 10^{-5}}$	$0 \times$
q_2: VB	0	0	$\mathbf{0{,}53\times 10^{-4}}$	0	$\mathbf{0{,}179\times 10^{-9}}$
q_1: PPSS	0	**0,025**	0	0	$0\times$
q_0: Start	**1**				
	o1:« s »	o2: I	o3:ωant	o4:to	O5: race

3. Terminaison

$$\hat{P} = \max_{1 \le i \le N} \delta_T(i)$$

$$\widehat{S_T} = \arg \max_{1 \le i \le N} \delta_T(i)$$

$$\boldsymbol{P} = \mathbf{0{,}179 \times 10^{-9}}$$

S5 = VB

Le chemin optimal est déterminé par un retour en arrière

$\widehat{S_T}, \psi_T(\widehat{S_T}), \psi_{T-2}(\widehat{S_{T-1}}), \ldots$

VB , To, VB, PPSS, « s » ,Null