

# Entropizer 1.1 : un outil informatique pour l'analyse séquentielle<sup>1</sup>

**Abstract**

**Résumé**

**Mots-clés:**

## **1. Introduction**

une chais s e ier prit

---

## 2. Les bases de l'analyse séquentielle

### 2.1 Les chaînes de Markov

$a_1 \ a_2 \ \dots \ a_m$

$X_t$

$P_{a_i \rightarrow a_j}$

$X_{t+1} = a_j | X_t = a_i$

$$p(a_i \rightarrow a_j) = P(X_{t+1} = a_j | X_t = a_i)$$

$P_{ij}$

$\omega_i \in A^k$

$P_{ij}$

### 2.2 Estimation des paramètres du modèle

$n$

$n \ a$

$k$

$\omega \in A^k$

$a \in A^k$

$f(a)$

$n \ a$

$$f(\omega) = \frac{n(\omega)}{n-k+1}$$

$p(\omega \rightarrow a)$

$a \in A$

$$p(\omega \rightarrow a) = \frac{n(\omega * a)}{\sum_{a \in A} n(\omega * a)}$$

### 2.3 Le concept d'entropie

$k$

$\omega \in A^k$

$X$

$H_k = -\sum_{\omega \in A^k} P(\omega) \log P(\omega)$

$H_k$

$A^k$

$h_k$

$H_k$

$H_{k-1}$

$k \geq 1$

$d_k \quad h_k \quad h_{k+1} \quad k \geq \quad k$

**2.4 Le test de l'ordre du processus**

$X \quad k$   
 $A^k \quad A^k \quad m^k$   
 $H_k \quad m^k \quad k$   
 $h_k \quad m \quad k$   
 $d_k \quad k$   
 $H_r \quad h_k \quad d_k \quad r \quad H_k \quad k \geq r \quad k \quad r \quad r \quad H_r \quad H_r \quad r$   
 $d_k$   
 $d_k \quad k \quad \alpha$   
 $H_0 \quad k \quad k$   
 $H_1 \quad k \quad k$   
 $H_0 \quad k \quad \alpha$   
 $(n-k)d_{k+} \geq \chi_{-\alpha} [m^k(m-)]$   
 $m \quad n \quad k$   
 $n \quad m \quad k$

**2.5 Exemple**

$n$   
 BACBCBCBAACBABCABCABCABCCBAABACCBAACABCABACBACCBACCBACBACBCABCABACBCCABCCAB  
 CCBACBAACBBCABCACBCABAABC  
 $n \quad m \quad k \quad k \leq \quad k$   
 $P_A \quad P_B \quad P_C$   
 $H_1 \quad H_3 \quad H_4 \quad h_1 \quad H_1 \quad h_2 \quad H_2 \quad H_1$   
 $H_2 \quad h_3 \quad h_4 \quad d_1 \quad h_1 \quad h_2 \quad d_2 \quad d_3$   
 $\alpha$   
 $k \quad (n-k)d_{k+} \quad \chi_{-\alpha} [m^k(m-)]$

$$\begin{array}{ccccccc}
 & & k & & & & H_0 k \\
 & k & k & & & & \alpha \\
 & & & & H_1 1 & & \\
 P & & & & & & \\
 BA & BB & BC & CA & CB & CC & \\
 P_{ij} & p & \omega_i \rightarrow a_j & a_j \in A & A B C & \omega_i \in A & AA AB AC
 \end{array}$$

$$P \left( \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array} \right)$$

### 3. Présentation du logiciel

#### 3.1 Généralités

$$H_k^\alpha = \frac{v_k}{- \alpha} \sum_{\omega \in A^k} P(\omega)^\alpha$$

#### 3.2 Distributions de k-grammes et probabilités de transition

$$\begin{array}{ccccccc}
 & & & & & & k \quad k' \\
 & & & & & & \omega \in A^k \\
 & & & & k' & & \\
 & & n \omega & & f \omega & & \\
 & & & & & & \\
 k & & k \quad k' & & & & \\
 & & & & a \in A & & \\
 & & & & & & \\
 & & & & A^k & & k \\
 n \quad k & & n & & & & k
 \end{array}$$

\_ p d → \_  
 d  
 d  
 Ulysses

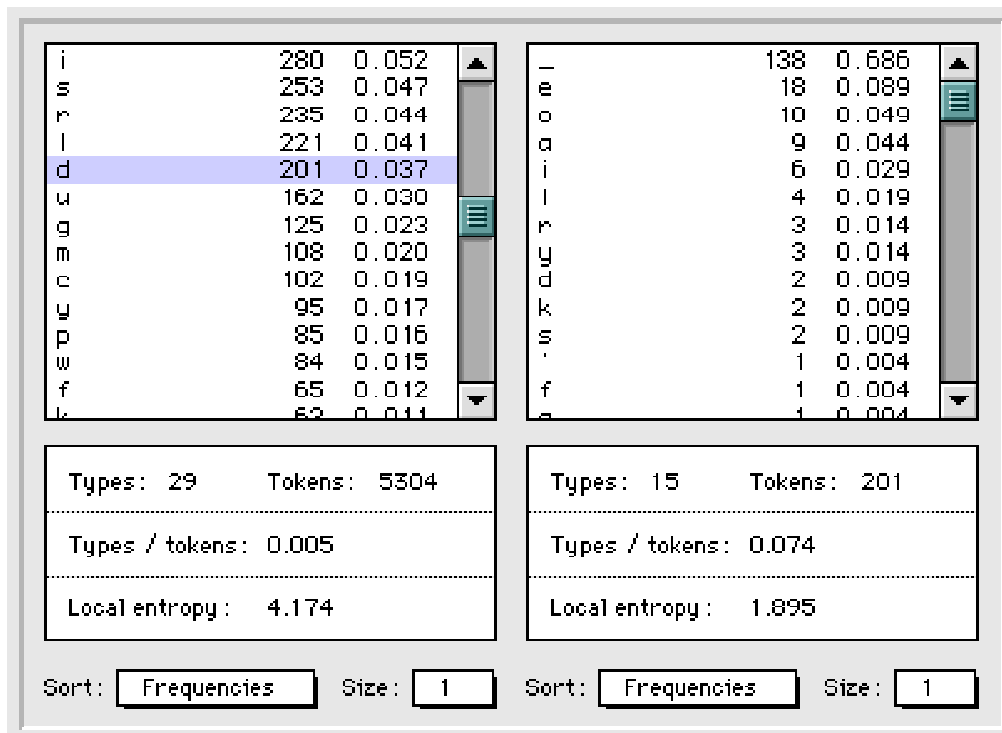


Figure 1: affichage des distributions de k-grammes

### 3.3 Entropies et test de l'ordre du processus

$k$

$\alpha$

$d_k$   $\alpha$

$d_2$   $k$

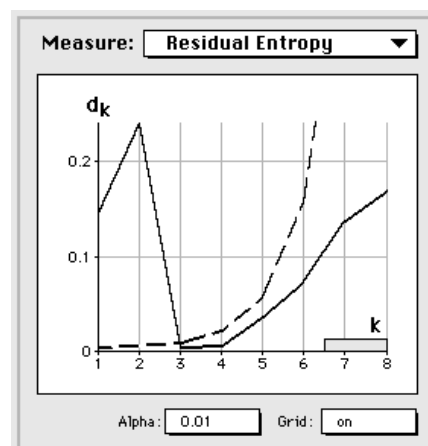
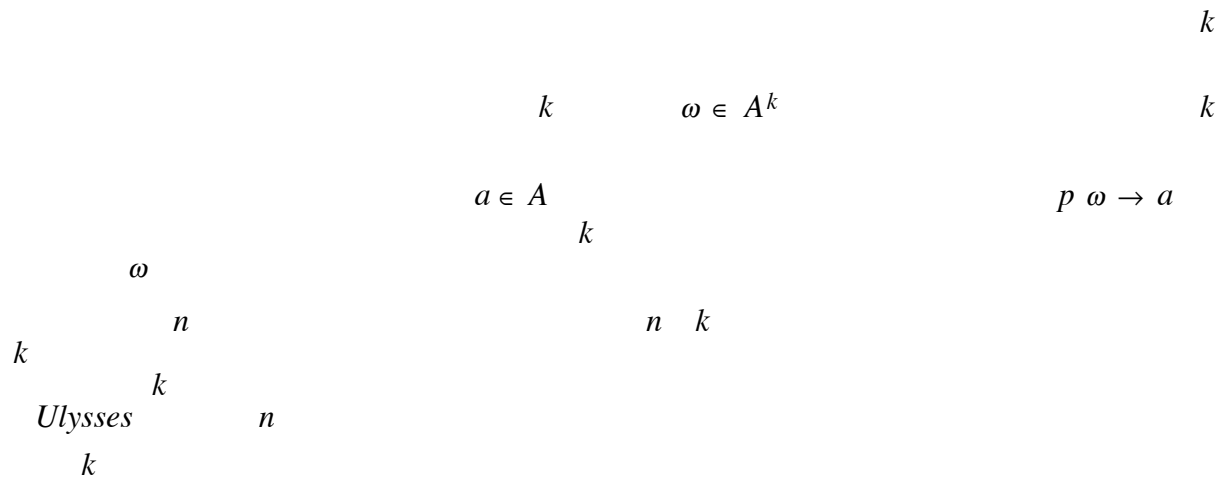


Figure 2: affichage de l'entropie résiduelle et test de l'ordre du processus

### 3.4 Simulation du processus



ene o y m doninocan's han lickndas s icken ofrd s fed teved f ind let clly  
 g s owighe y celutowacke whegint thearathe ed aisoridimucond myod  
 leensetthe iere

id somin broming ton we an liganch the coure raying bo yours a cardeard did  
 cou by an oned iffew i do sing swither car a ving sin sphent ing eten bried  
 buts sho

he towards head boreadful acrose rounger from it up he dang voice ans -- a  
 spoke blackerchs abouth remembermalach he door he what the lover a secread  
 had to his

### 3.5 Exemple d'application

[...]rtes de phrases ? L'affirmation, l'interrogation, le commandement peut-  
 être ? - Il en est d'in[...]

[...]CCVC CV CCCVCVC CVCCVCCVCVVC CVCCVCCVCVVC CV CVCCVCCVCVCCVVC VCCV VC  
 VC VCC CVC[...]

---

y w n m

$$k \leq k \leq$$

$$P \left( \begin{array}{c} \\ \\ \\ \end{array} \right)$$

$$P_{ij} = p(a_i \rightarrow a_j) \quad a_i, a_j \in A \quad \_ \text{ c v}$$

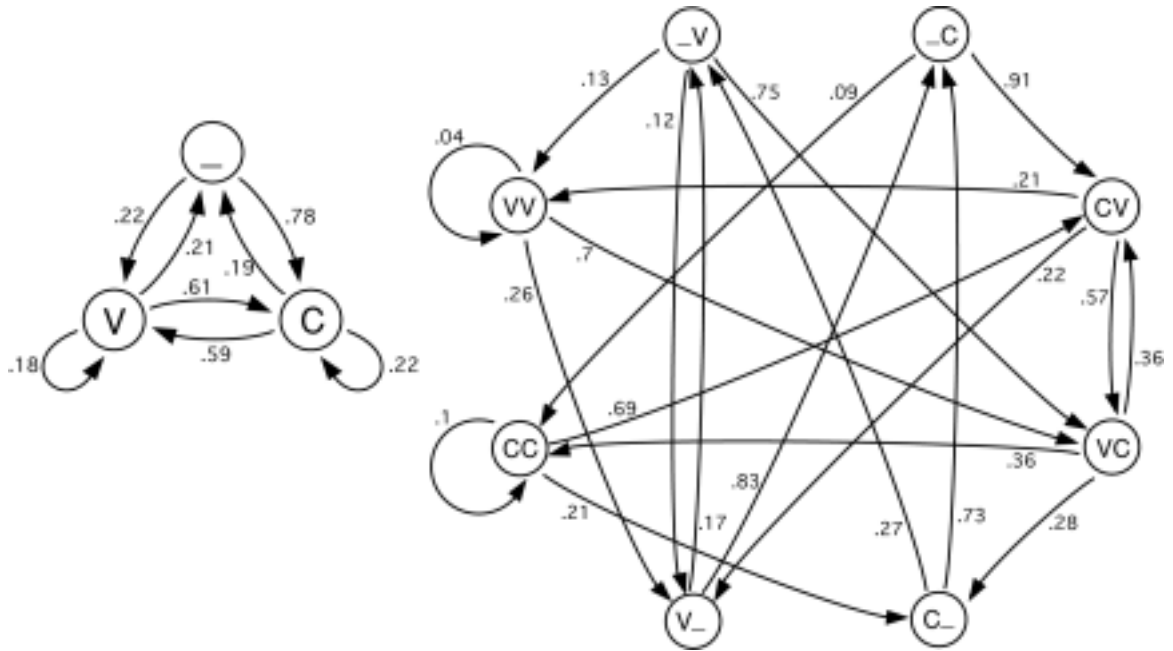


Figure 3: chaînes de Markov d'ordre 1 et 2 pour les voyelles et consonnes du Français

$k$	$\omega_k$	$p(\omega_k \rightarrow \_)$	$p(\omega_k \rightarrow c)$	$p(\omega_k \rightarrow v)$
	$c$	---	---	---
	$vc$	---	---	---
	$cvc$	---	---	---
	$\_cvc$	---	---	---
	$c\_cvc$	---	---	---
	$cc\_cvc$	---	---	---
	$vcc\_cvc$	---	---	---
	$k$			

$$k \geq$$

$m^k$   $m$   $k$

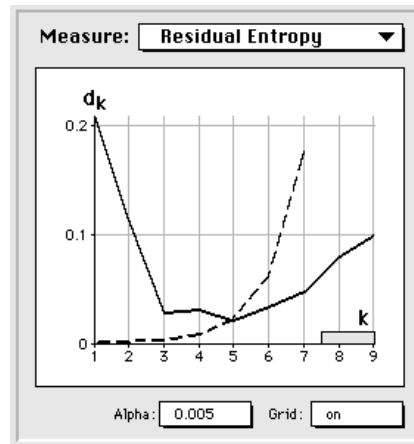


Figure 4: entropie résiduelle pour les voyelles et consonnes du Français

#### 4. Conclusion

*ad hoc*

#### Références

*Modèles et Données*

*Proceedings of the  
5th International Conference on the Statistical Analysis of Textual Data (JADT 2000)*

*Markov Models and Linguistic Theory  
Sequential Analysis*

*Structural Linguistics*

*Bell Syst. tech. Journal*

*Codes and Cryptography  
Entropizer 1.1*