LABORATOIRE
D'INFORMATIQUE
& SYSTÈMES
UMR 7020

UNIVERSITÉ DE
TOULON

Théorie de l'Information

Chaîne de communication

Julien SEINTURIER
Maître de Conférences

Licence d'Informatique 3ieme année – 2022 / 2023

Intérêt de l'information

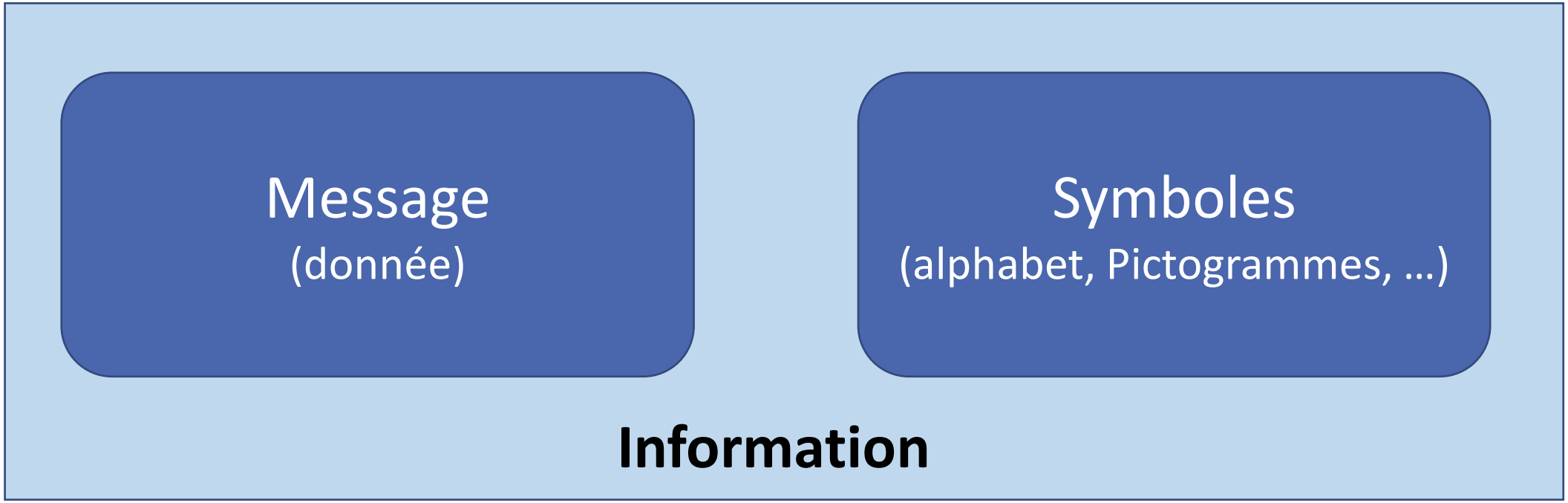
- Faire évoluer la connaissance
- Prendre des décisions
- **Partager**

Pas d'information sans communication

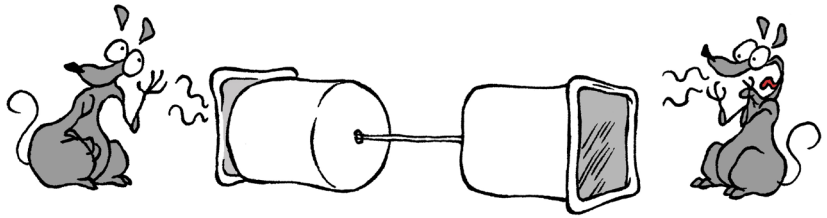
https://journals.openedition.org/questionsdecommunication/254

Information

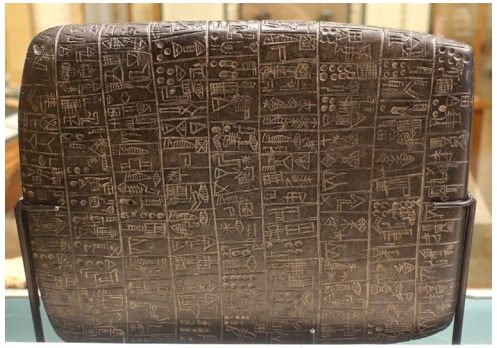
Définition: En informatique et en télécommunication, l'information est un élément de connaissance (donnée) susceptible d'être conservé, traité ou transmis via un **signal** sous forme de **message** codé à l'aide de **symboles** .



Communication



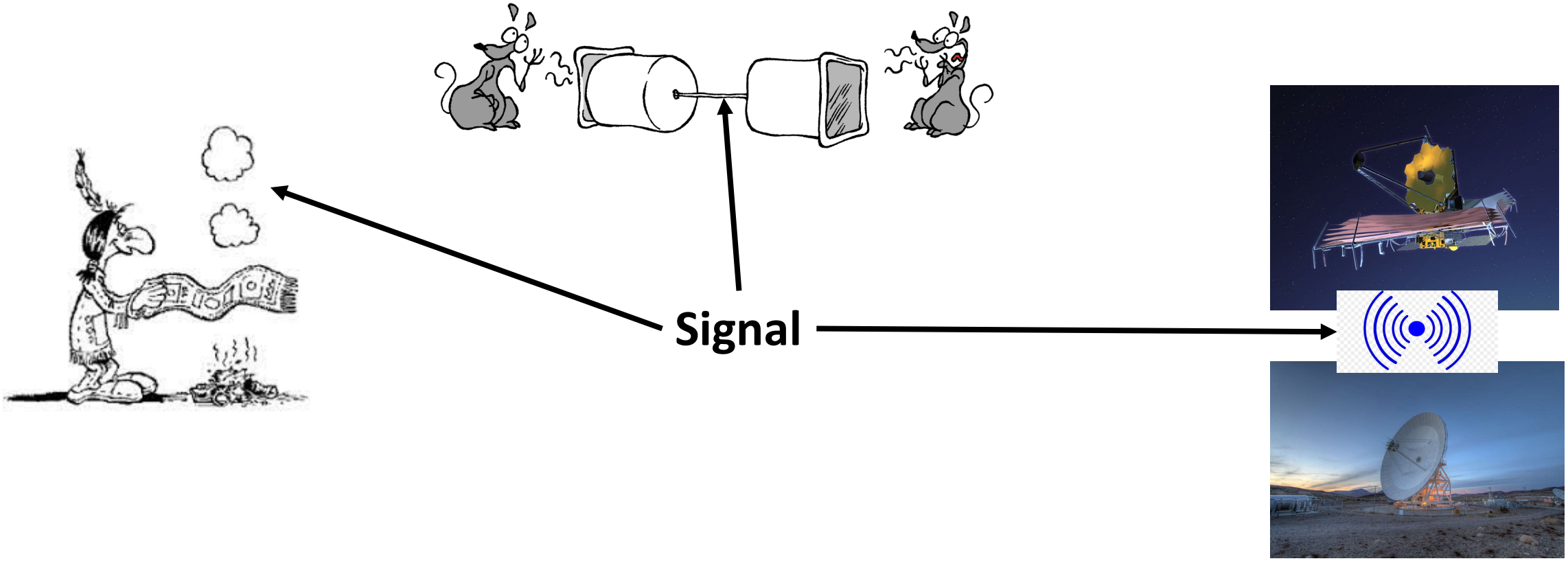
Yaourtphone
© espace des Sciences (espace-sciences.org)



James Webb Spatial Telescope (JWST)
© espace des Sciences (espace-sciences.org)

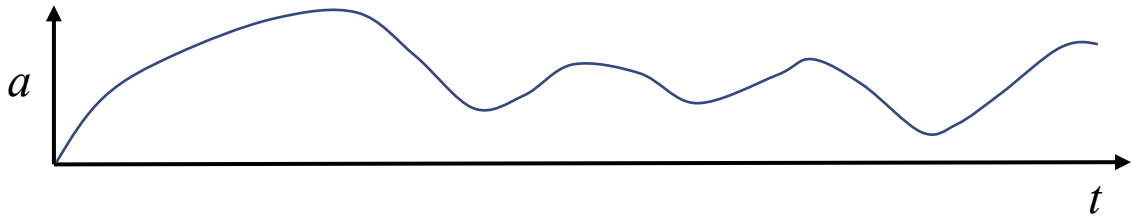
Signal

Définition: On appelle signal est une **quantité mesurable** porteuse **d'information**. On peut classer les signaux par leur usage, le type de message qu'ils portent ou le moyen de transmission.

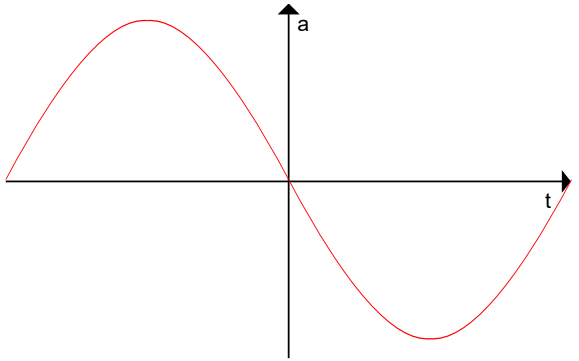


Signal Analogique

Définition: Un signal analogique est un signal qui prend une **infinité de valeurs** qui varient en **continu** dans **le temps**. Par convention on note $a(t)$ l'**amplitude** du signal s à l'instant t.



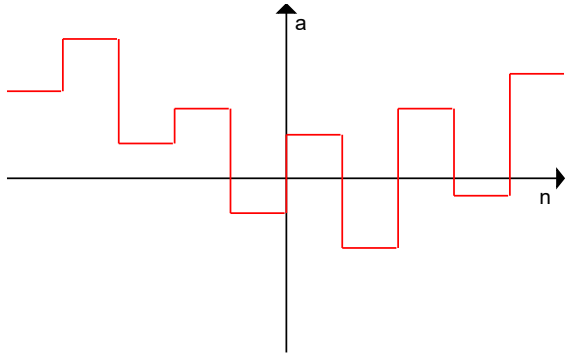
Mathématiquement, un signal analogique est représenté par une fonction continue (le plus souvent sur $\mathbb{R}$).
Parmi les famille de signaux analogiques, une des plus connue est celle des signaux sinusoïdaux.



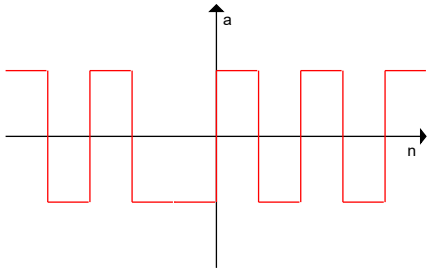
Exemple: Tension électrique, intensité électrique, son, température, VGA, …

Signal Numérique

Définition: Un signal numérique est un signal qui prend un **nombre fini de valeurs** évoluant selon une variable **discrète** dont les valeurs sont appelées **échantillons**. Par convention on note $a[n]$ l'**amplitude** du signal s pour l'échantillon n.



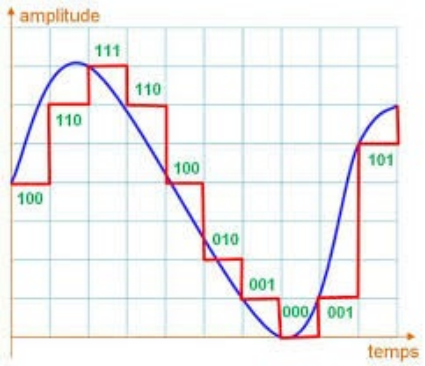
Parmi les signaux numériques les plus connus, le signal binaire:



Exemple: RJ45, fibre optique, BUS informatique, HDMI, …

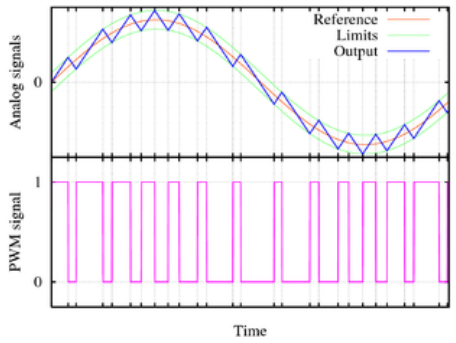
Passage Analogique / Numérique

- Analogique vers numérique: Echantillonnage

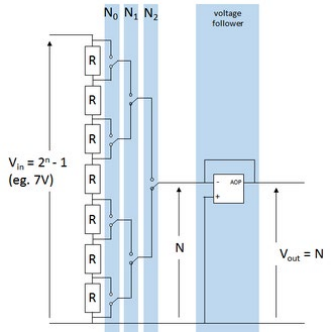


Convertisseur analogique Numérique (**CAN**)

- Numérique vers analogique



Modulation

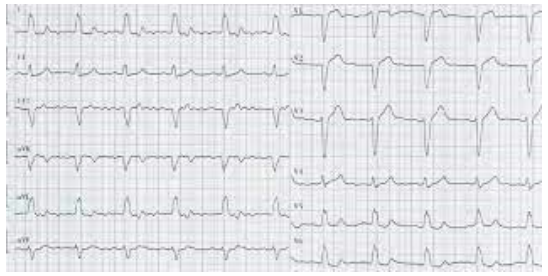


Resistance

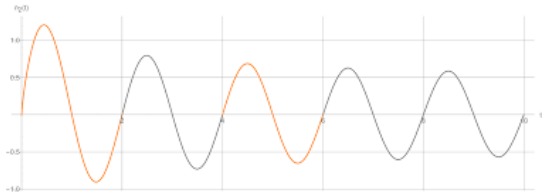
Convertisseur numérique analogique (**CNA**)

Problèmes potentiels

- Altération
  - Parasites
  - Atténuation
- Raisons
  - Matériel défectueux
  - Milieu
  - Distance

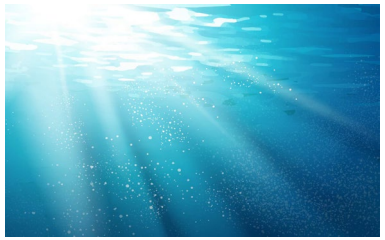


ECG parasité



Signal atténué

Câble endommagé



Wifi Vs Eau

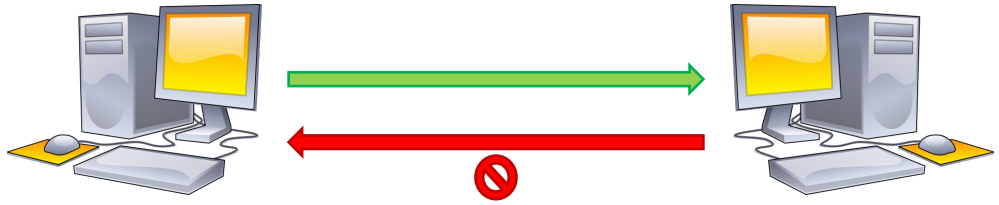
Distances astronomiques

Simplex

La communication ne s'établit que dans **un seul sens**.
Ex: Caméra, Streaming

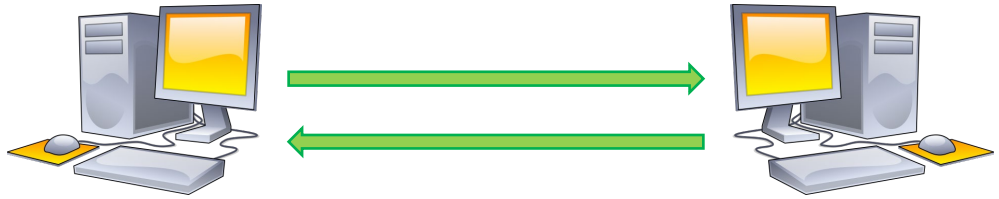
Half duplex

La communication ne s'établit que dans **un seul sens à la fois**.
Ex: Talkie Walkie, Radio



Full duplex

La communication ne s'établit dans **les deux sens simultanément**.
Ex: Téléphone, Ethernet



Problèmes potentiels

- Mauvaise activation (half-duplex)

  Envoi non accepté ou interrompu (oubli d'appuie sur le bouton du Talkie Walkie)

- Interaction entre communications simultanées (full-duplex)

  Personnes parlant en même temps au téléphone, collision de paquets
  Ethernet, …

Point de vue théorique

- communication half-huplex = 2 communications simplex alternatives
- communication full-huplex = 2 communications simplex simultanées

Formalisation de Shannon

La **chaîne de communication** numérique représente les différentes **étapes** de **traitement** de l'**information**. Elle relie un **émetteur** à un **récepteur** par l'intermédiaire d'un **canal de transmission**. Le canal est le milieu dans lequel est **transmis** ou **stocké** l'**information** sous forme de **signal**.

Formalisée en **1949** par Claude E. Shannon dans *The Mathematical Theory of Communication*

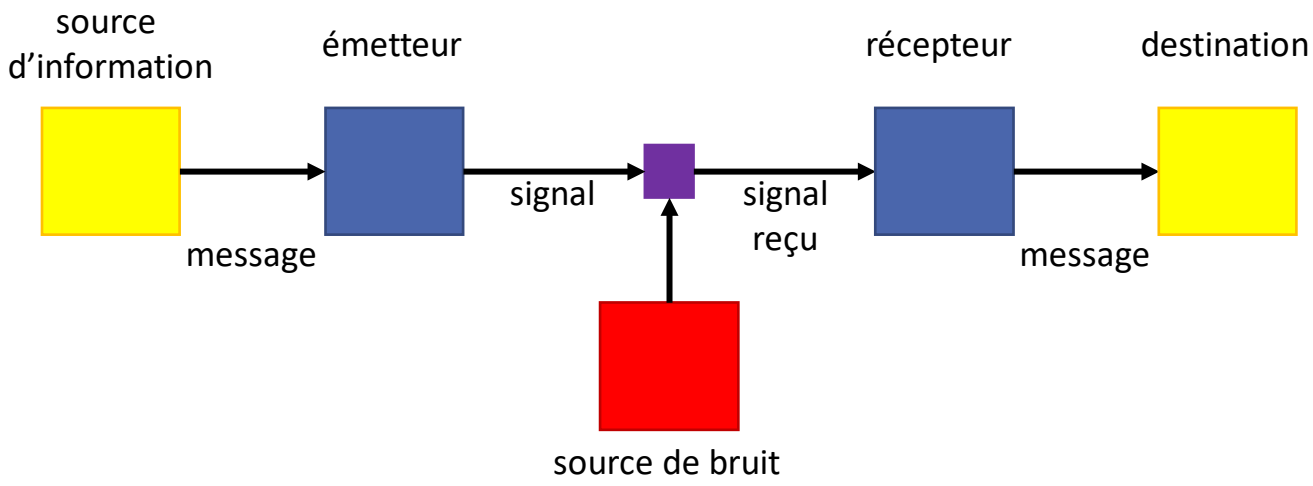


Schéma d'un système général de communication (Shannon, 1949)

Formalisation de Shannon

- Tout problème de transmission de l'information résulte d'un un bruit
- Transmission de l'Information seulement dépendante:
  - Du canal de transmission
  - Du bruit

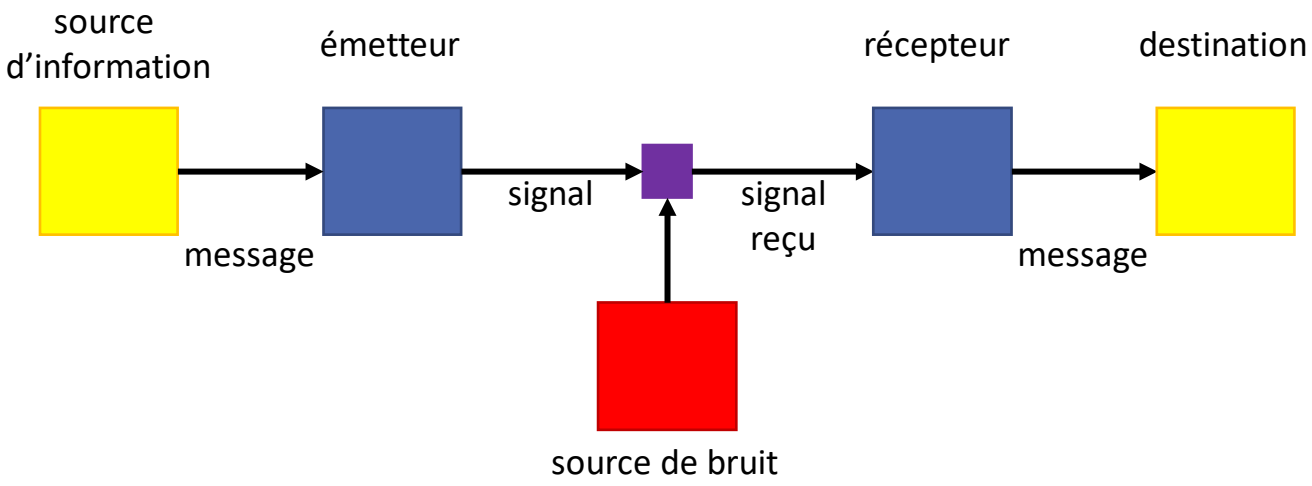
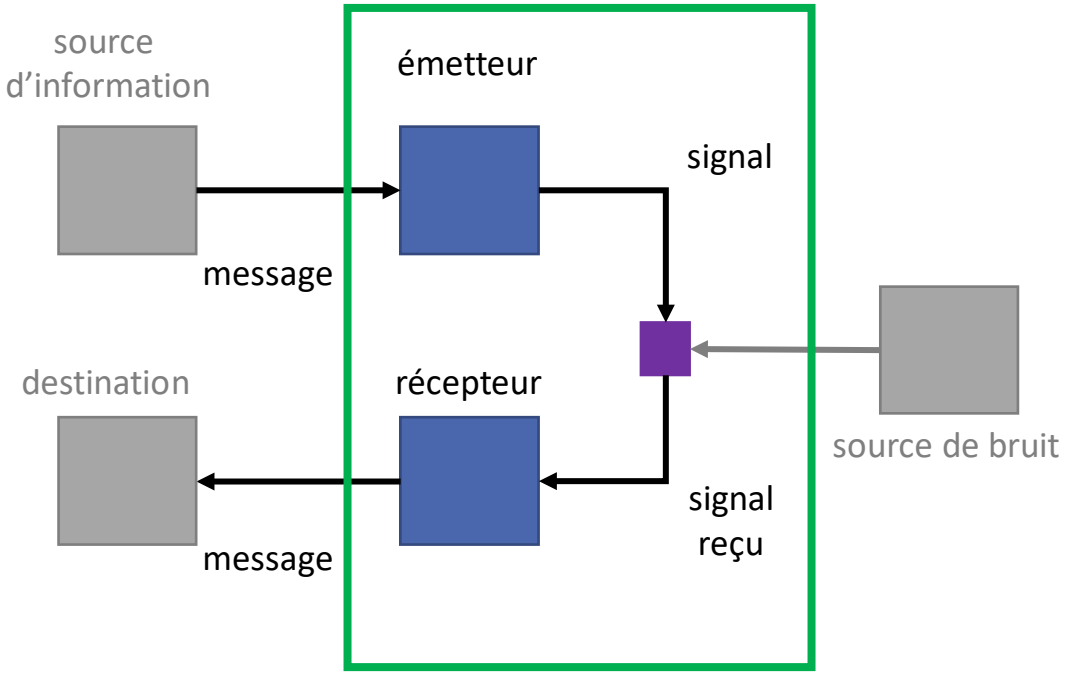


Schéma d'un système général de communication (Shannon, 1949)

Emission / réception

- Emetteur et récepteur réalisent plusieurs opérations critiques



- Etudier ces différences opérations

Codage de source

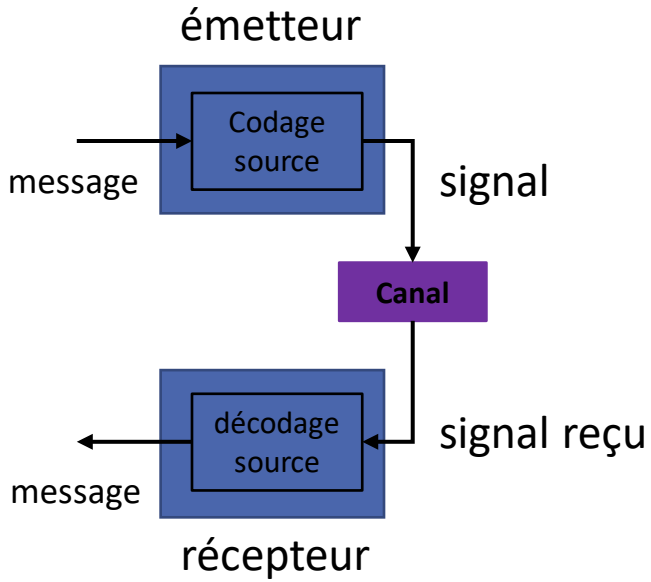
- Détermine la forme du message (son codage)

  Ex: des nombres décimaux (valeur de températures)

- Choix de l'ensemble des symboles utilisés

  Ex: nombres décimaux entre -100 et 100 avec 3 décimales

- Quantifier l'information à transmettre
  - Anticiper le choix du canal / du signal
  - Evaluer si le canal : signal choisis sont adéquat

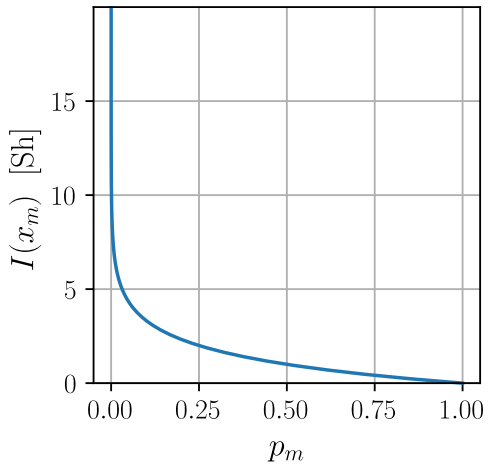


Auto-information

Définition: Soit $S = \{s_1, \ldots, s_n\}$ l'ensemble des symboles sur lesquels une source information X est codée et soit p_m la probabilité d'apparition du symbole s_m. L'auto-information $I(s_m)$ associée au symbole s_m est telle que:

- Une source produisant toujours le même symbole n'apporte aucune information: $I(s_m) = 0$ si $p_m = 1$
- Un symbole n'apparaissant jamais a potentiellement une information maximale: $I(s_m) \longrightarrow +\infty$ si $p_m = 0$
- l'information portée par deux symboles est la même que deux informations: $I(s_i s_j) = I(s_i) + I(s_j)$

La définition $\boldsymbol{I(s_m) = -\log_2(p_m)}$ respecte les 3 postulats précédents.



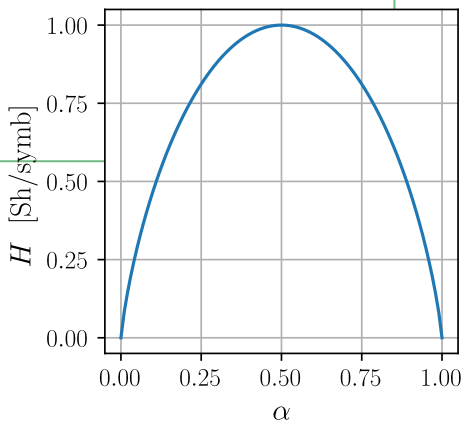
Lettre	Anglais	Français
a	0,08167	0.07636
b	0,01492	0.00901
c	0,02782	0,03260

Entropie

Définition: L'entropie d'une source X codée sur un ensemble de symboles $S = \{s_1, \ldots, s_n\}$ est la somme des auto-informations de chaque symbole s_i pondéré par sa probabilité d'apparition p_i. Formellement:

$$H(X) = \sum_{i=1}^{n} p_i \times I(s_i) = -\sum_{i=1}^{n} p_i \times log_2(p_i)$$

L'entropie est exprimée en Shannon par symbole ($Sh \cdot symb^{-1}$)



- L'entropie est telle que $0 \leq H(X) \leq log_2(n)$
  - Si X émet toujours le même symbole: $H(X) = 0$
  - Si tous les symboles émis par X ont la même probabilité d'apparition:

$$H(X) = \sum_{i=1}^{n} \frac{1}{n} \times I(s_i) = -\sum_{i=1}^{n} \frac{1}{n} \times log_2\left(\frac{1}{n}\right) = -\sum_{i=1}^{n} -\frac{log_2(n)}{n} = log_2(n)$$

Entropie

Exemple:

On souhaite calculer l'entropie moyenne d'une source X composée d'un texte rédigé en français.

$S = \{ A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z \}$

Entropie

Exemple:

On souhaite calculer l'entropie moyenne d'une source X composée d'un texte rédigé en français.

$S = \{ A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z \}$

A	0.0871	**G**	0.0097	**M**	0.0323	**S**	0.0791	**Y**	0.0019
B	0.0094	**H**	0.0108	**N**	0.0642	**T**	0.0711	**Z**	0.0021
C	0.0316	**I**	0.0699	**O**	0.0536	**U**	0.0615		
D	0.0355	**J**	0.0072	**P**	0.0304	**V**	0.0184		
E	0.1784	**K**	0.0016	**Q**	0.0089	**W**	0.0004		
F	0.0096	**L**	0.0568	**R**	0.0643	**X**	0.0042		

Probabilité d'apparition des lettres en langue française écrite
Source bepo.fr

Entropie

Exemple:

On souhaite calculer l'entropie moyenne d'une source X composée d'un texte rédigé en français.

$S = \{ A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z \}$

A	0.0871	**G**	0.0097	**M**	0.0323	**S**	0.0791	**Y**	0.0019
B	0.0094	**H**	0.0108	**N**	0.0642	**T**	0.0711	**Z**	0.0021
C	0.0316	**I**	0.0699	**O**	0.0536	**U**	0.0615		
D	0.0355	**J**	0.0072	**P**	0.0304	**V**	0.0184		
E	0.1784	**K**	0.0016	**Q**	0.0089	**W**	0.0004		
F	0.0096	**L**	0.0568	**R**	0.0643	**X**	0.0042		

Probabilité d'apparition des lettres en langue française écrite
Source bepo.fr

$$H(X) = \sum_{i=1}^{n} p_i \times I(s_i) = 0.0871 \times log_2(0.0871) + \cdots + 0{,}0021 \times log_2(0{,}0021) = \mathbf{4.0080}$$

Code source

Définition: Soit une source d'information codée sur un ensemble de symbole $S = \{s_1, \ldots, s_n\}$ et soit un canal utilisant un alphabet $A = \{a_1, \ldots, a_m\}$. Un **code source** associe à chaque symbole s_i un **mot** composé d'une suite de l_i symboles issus de A.

Code source

Définition: Soit une source d'information codée sur un ensemble de symbole $S = \{s_1, \ldots, s_n\}$ et soit un canal utilisant un alphabet $A = \{a_1, \ldots, a_m\}$. Un **code source** associe à chaque symbole s_i un **mot** composé d'une suite de l_i symboles issus de A.

Exemple:

On souhaite coder de l'information exprimée en alphabet latin sur un canal utilisant le Morse (télégraphe). La code source choisi associe à chaque lettre sa représentation en Morse.

$S = \{ A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z \}$

A = {●, ▬}, symboles des deux signes du morse utilisés par le canal

Le symbole « F », $6^{ième}$ élément de S est codé en morse par le mot de 4 symboles « ● ● ▬ ● »
On a donc $l_6 = 4$.

Longueur de code

Définition: Soit une source d'information codée sur un ensemble de symbole $S = \{s_1, \ldots, s_n\}$ pour lequel chaque symbole s_i à une probabilité d'apparition p_i, soit un canal utilisant un alphabet $A = \{a_1, \ldots, a_m\}$ et soit un code source associant à chaque symbole s_i de S un mot de longueur l_i. La longueur moyenne du code, notée L, est définie par:

$$L = \sum_{i=1}^{n} p_i \times l_i$$

Longueur de code

Définition: Soit une source d'information codée sur un ensemble de symbole $S = \{s_1, \ldots, s_n\}$ pour lequel chaque symbole s_i à une probabilité d'apparition p_i, soit un canal utilisant un alphabet $A = \{a_1, \ldots, a_m\}$ et soit un code source associant à chaque symbole s_i de S un mot de longueur l_i. La longueur moyenne du code, notée L, est définie par:

$$L = \sum_{i=1}^{n} p_i \times l_i$$

Exemple: Calculer la longueur moyenne de code pour un codage en morse de textes écrits en français.

A	0.0871	G	0.0097	M	0.0323	S	0.0791	Y	0.0019
B	0.0094	H	0.0108	N	0.0642	T	0.0711	Z	0.0021
C	0.0316	I	0.0699	O	0.0536	U	0.0615		
D	0.0355	J	0.0072	P	0.0304	V	0.0184		
E	0.1784	K	0.0016	Q	0.0089	W	0.0004		
F	0.0096	L	0.0568	R	0.0643	X	0.0042		

Probabilité d'apparition des lettres en langue française écrite
Source bepo.fr

A	•—	G	——•	M	——	S	•••	Y	—•——
B	—•••	H	••••	N	—•	T	—	Z	——••
C	—•—•	I	••	O	———	U	••—		
D	—••	J	•———	P	•——•	V	•••—		
E	•	K	—•—	Q	——•—	W	•——		
F	••—•	L	•—••	R	•—•	X	—••—		

Code Morse international, source UIT

$$L = \sum_{i=1}^{n} p_i \times l_i = 0.0871 \times 2 + \cdots + 0.0021 \times 4 = \mathbf{2.4388}$$

Théorème du codage de source

Soit X une source d'information d'entropie $H(X)$ et soit un code source de longueur moyenne L, Une transmission de l'information avec un taux d'erreur le plus petit possible peut être réalisée si $L \geq H(X)$

Premier théorème de Shannon

Théorème du codage de source

Soit X une source d'information d'entropie $H(X)$ et soit un code source de longueur moyenne L, Une transmission de l'information avec un taux d'erreur le plus petit possible peut être réalisée si $L \geq H(X)$

Premier théorème de Shannon

Exemple:

Peut-on transmettre sur un canal en morse les informations d'une source X composée d'un texte rédigé en français ?

$S = \{ A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z \}$

$A = \{\bullet, \text{—}\}$

$$H(X) = \sum_{i=1}^{n} p_i \times I(s_i) = \mathbf{4.0080} \qquad L = \sum_{i=1}^{n} p_i \times l_i = \mathbf{2.4388}$$

$L < H(X)$, la transmission est impossible sans erreur.

Débit de source

Définition: Le **débit source**, noté D_s, est le nombre de symboles émis par la source en une seconde. Il s'exprime en $symb \cdot s^{-1}$.

Débit de source

Définition: Le **débit source**, noté D_s, est le nombre de symboles émis par la source en une seconde. Il s'exprime en $symb \cdot s^{-1}$.

Taux d'émission

Définition: Soit X une source d'information d'entropie $H(X)$ et de débit de source D_s. La quantité d'information produite par cette source, appelée **taux d'émission** et notée T_s est définie par:

$$T_S = H(X) \times D_s$$

Le taux d'émission s'exprime en $Sh \cdot s^{-1}$

Débit canal

Définition: Le débit canal, noté D_c est le nombre de symboles transmis par le canal en une seconde. Il peut être égal au débit de source mais il peut également être différents si les symboles utilisés par le canal sont différents de ceux utilisés pour coder la source.

$S = \{ A, \ldots, Z \}$ Symboles information (26 lettres)

$A = \{\cdot, -\}$ Alphabet canal (Morse, 2 symboles)

Débit canal

Définition: Le débit canal, noté D_c est le nombre de symboles transmis par le canal en une seconde. Il peut être égal au débit de source mais il peut également être différents si les symboles utilisés par le canal sont différents de ceux utilisés pour coder la source.

$S = \{ A, \ldots, Z \}$ Symboles information (26 lettres)

$A = \{\cdot, -\}$ Alphabet canal (Morse, 2 symboles)

Capacité de canal

Définition: Soit un canal dont le codage se fait sur N symboles et soit D_c le débit du canal, la capacité de canal, notée C_c est définie telle que:

$$C_c = D_c \times log_2(N)$$

La capacité s'exprime en $Sh \cdot s^{-1}$

Théorème du codage canal

Soit un canal de capacité C_c, il est possible de transmettre sans erreur des informations issue d'une source de taux d'émission T_s si et seulement si $C_c \geq T_s$

Deuxième théorème de Shannon

Théorème du codage canal

Soit un canal de capacité C_c, il est possible de transmettre sans erreur des informations issue d'une source de taux d'émission T_s si et seulement si $C_c \geq T_s$

Deuxième théorème de Shannon

Ce théorème repose uniquement sur la capacité d'un signal, les symboles et alphabets utilisés pour coder l'information et le débit de la source.

Le codage utilisé n'a aucune influence sur son résultat.

Exercice

Soit X une source d'information codée sur l'ensemble de symboles $S = \{A, B, C\}$ dont les probabilités respectives sont 0.8, 0.15 et 0.05. Le débit de cette source est de 100 symboles par secondes. La transmission d'information se fait via un canal binaire d'alphabet $A = \{0, 1\}$ pouvant transmettre 95 bits par seconde sans erreur.

1. Calculer l'auto information des symboles.
2. Calculer l'entropie de la source.
3. Quel est le débit de source ?
4. Calculer le taux d'émission de la source
5. Quel est le débit du canal ?
6. Quelle est la capacité du canal ?
7. Le canal permet-il de transmettre les informations de la source.
8. Soit le code source suivant: $A \rightarrow 0$, $B \rightarrow 1$, $C \rightarrow 01$. Celui-ci permet-il d'encoder la source d'information X sans erreur ?

Exercice

1. Calculer l'auto information des symboles.

$$I(S_1) = I(A) = -\log_2(0.80) \approx 0.32$$
$$I(S_2) = I(B) = -\log_2(0.15) \approx 2.74$$
$$I(S_3) = I(C) = -\log_2(0.05) \approx 4.32$$

2. Calculer l'entropie de la source.

$$H(X) = \sum_{i=1}^{3} p_i \times I(s_i) = 0.80 \times 0.32 + 0.15 \times 2.74 + 0.05 \times 4.32 = 0.26 + 0.41 + 0.22 = \mathbf{0.88\ Sh \cdot symb^{-1}}$$

3. Quel est de débit de la source ?

Le débit de la source, noté D_s , est de $\mathbf{100\ symb \cdot s^{-1}}$.

4. Calculer le taux d'émission de la source.

$$T_S = H(X) \times D_S = 0{,}88 \times 100 = \mathbf{88\ Sh \cdot s^{-1}}$$

Exercice

5. Quel est le débit du canal.

Le débit du canal, noté D_c, est de $\mathbf{95\ symb \cdot s^{-1}}$.

6. Quelle est la capacité du canal.

$$C_c = D_C \times log_2\,(N) = D_C \times log_2\,(2) = 95 \times 1 = \mathbf{95\ Sh \cdot s^{-1}}$$

7. Le canal permet-il de transmettre les informations de la source ?

D'après le théorème du codage de canal (second théorème de Shannon), un canal de capacité C_c, peut transmettre sans erreur des informations issue d'une source X de taux d'émission T_s si et seulement si $C_c \geq T_s$.

Dans notre cas, $C_C = 95\ Sh{\cdot}s^{-1}$ et $T_s = 88\ Sh{\cdot}s^{-1}$. On a donc bien $95 \geq 88$, soit $\mathrm{C_C} \geq \mathrm{T_S}$. La transmission est possible sans erreur.

Exercice

8. Soit le code source suivant: $A \rightarrow 0, B \rightarrow 1, C \rightarrow 01$. Celui-ci permet-il d'encoder la source d'information X sans erreur ?

S	p_i	l_i
A	0.80	1
B	0.15	1
C	0.05	2

Les longueurs des mots codés sont: $l_1 = l_A = 1$, $l_2 = l_B = 1$ et $l_3 = l_C = 2$.

La longueur moyenne du code, notée L est donc:

$$L = \sum_{i=1}^{3} p_i \times l_i = 0.80 \times 1 + 0.15 \times 1 + 0.05 \times 2 = 0.80 + 0.15 + 0.10 = \mathbf{1.05}$$

De plus, d'après la question 2, l'entropie de X est $H(X) = \mathbf{0.88}$.

D'après le théorème du codage de source (premier théorème de Shannon), la transmission d'une source d'information X d'entropie $H(X)$ en utilisant un code source de longueur moyenne L avec un taux d'erreur le plus petit possible peut être réalisée si $L \geq H(X)$.

Dans notre cas $H(X) = 0.88$ et $L = 1.05$. On vérifie bien que $1.05 \geq 0.88$, la transmission peut se faire avec le code choisi.

Propriétés des codes sources

- Un code de **longueur fixe** est un code dont tous les mots ont le même nombre de symboles. Dans le cas contraire le code est de **longueur variable**.

  Exemple: Le code qui associe les symboles A à 00, B à 01 et C à 10 est de longueur fixe. Le code qui associe le symbole A à 0, le symbole B à 1 et le symbole C à 10 est de longueur variable.

- Un code a **décodage unique** n'admet aucune ambiguïté lors de son décodage.

  Exemple: Le code Morse, sans ajouter de coupures entre les lettres, n'est pas à décodage unique. En effet: ●●●● peut être EEEE (● ● ● ●), II(●● ●●), EIE (● ●● ●), ES (● ●●●), SE (●●● ●), …

- Un **code instantané** est un code qui permet de décoder le symbole courant sans avoir à attendre le symbole suivant.

  Exemple: Le code qui associe le symbole A à 0 et le symbole B à 10 est instantané. Si l'on reçoit un 0 il s'agit forcement d'un A, si on reçoit un 1, il ne peut plus s'agir d'un A et donc forcement d'un B.

Optimiser les codes

- Contraintes de la théorie
  - Premier théorème de Shannon: $L \geq H(X)$
  - Second théorème de Shannon: $C_c \geq T_s$
- Solutions possibles
  - Diminuer la taille des mots du code source:
    trouver le plus petit L vérifiant $L \geq H(X)$
  - Utiliser à son avantage les distributions de probabilités
  - Principe de la compression

Codage de Huffman

- Proposé par David Huffman (1952)
  A method for the construction of minimum-redundancy codes

- Propriétés
  - Compression **sans perte** (conservation de l'entropie)
  - Code à **longueur variable**
  - Code à **décodage unique**
  - Utilise l'alphabet binaire $A = \{0, 1\}$

- **Code optimal** (il n'existe pas de code sans perte de longueur moyenne inférieure)

Codage de Huffman

Méthode

Soit une source d'information X codée sur un ensemble de symboles $S = \{s_1, \ldots, s_n\}$ et soit p_i la probabilité d'apparition du symbole s_i.

1. Ordonner les symboles de S dans l'ordre des probabilités décroissantes:

$$\forall i, j \in [1, n], s_i \leq s_j \leftrightarrow si\ p_i \leq p_j$$

2. Ajouter au code du symbole de plus faible probabilité, noté S_M, le bit 0 à gauche
3. Ajouter au code du deuxième symbole de plus faible probabilité, noté S_{M-1}, le bit 1 à gauche
4. Combiner S_M et S_{M-1} pour former un nouveau symbole de probabilité $p(S_M) + p(S_{M-1})$
5. Retourner à l'étape 2 tant qu'il reste plus d'un seul symbole

Codage de Huffman

Exemple: Coder le mot BONJOUR

1. Tri des symboles

R	N	U	O	B	J
0.0643	0.0642	0.0615	0.0536	0.0094	0.0072

2. Ajout de 0 à J

R	N	U	O	B	J
				1	0

3. Ajout de 1 à B

4. Combinaison de J et B

R	N	U	O	BJ
0.0643	0.0642	0.0615	0.0536	0.0166

Itération 2

2. Ajout de 0 à B et J

R	N	U	O	B	J
			1	01	00

3. Ajout de 1 à O

4. Combinaison de BJ et O

OBJ	R	N	U
0,0702	0.0643	0.0642	0.0615

Codage de Huffman

Itération 3

R	N	U	O	B	J
			1	01	00

OBJ	R	N	U
0,0702	0.0643	0.0642	0.0615

2. Ajout de 0 à U

R	N	U	O	B	J
	1	0	1	01	00

OBJ	R	N	U
0,0702	0.0643	0.0642	0.0615

3. Ajout de 1 à N

R	N	U	O	B	J
	1	0	1	01	00

NU	OBJ	R
0.1257	0,0702	0.0643

4. Combinaison de U et N

Itération 4

R	N	U	O	B	J
0	1	0	11	101	100

NU	OBJ	R
0.1257	0,0702	0.0643

2. Ajout de 0 à R

3. Ajout de 1 à O, B et J

R	N	U	O	B	J
0	1	0	11	101	100

ROBJ	NU
0,1345	0.1257

4. Combinaison de R et OBJ

Codage de Huffman

Itération 5

R	N	U	O	B	J	ROBJ	NU
0	1	0	11	101	100	0,1345	0.1257

2. Ajout de 0 à N et U

R	N	U	O	B	J	ROBJ	NU
10	01	00	111	1101	1100	0,1345	0.1257

3. Ajout de 1 à R, O, B et J

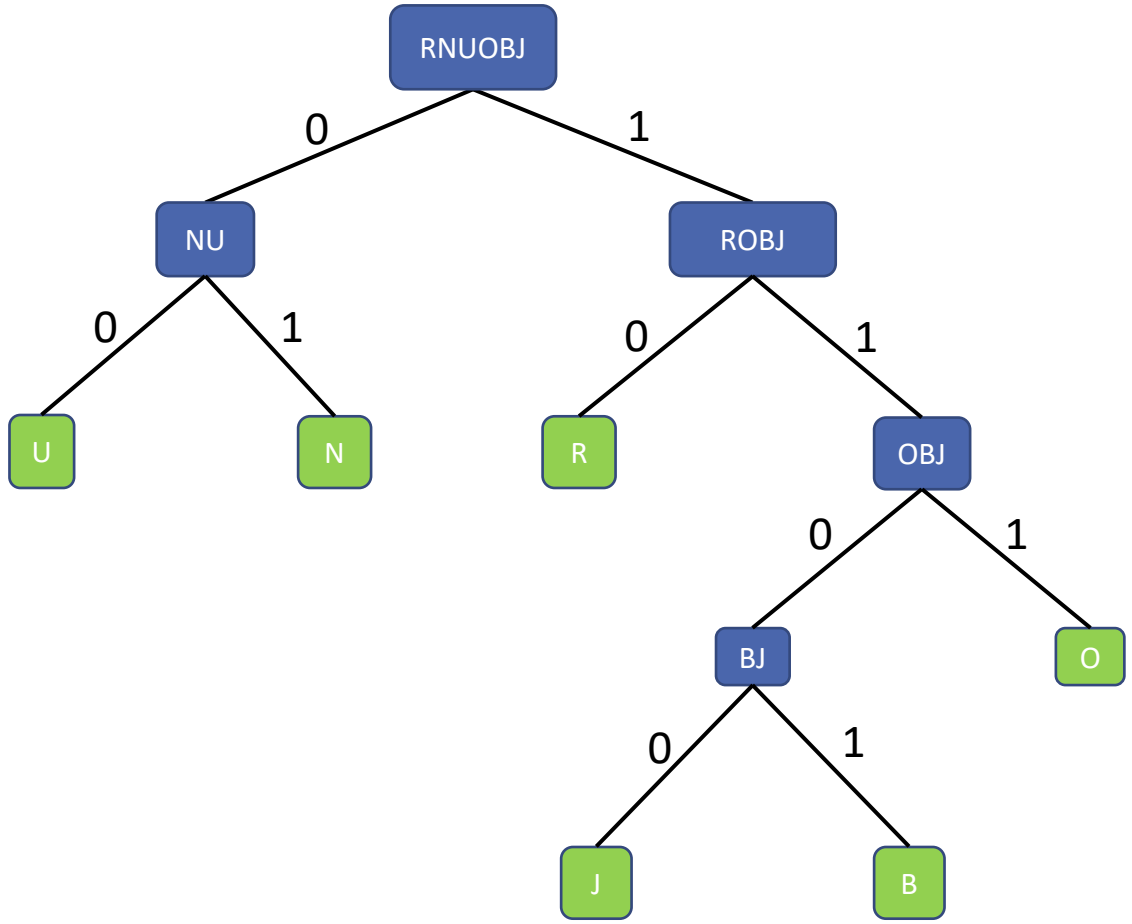
4. Combinaison de ROBJ et NU

R	N	U	O	B	J	ROBJNU
10	01	00	111	1101	1100	0,2602

Codage terminé

Codage de Huffman

- Equivalent à la construction d'un arbre binaire
- Les symboles en sont les feuilles
- Trouver un code = parcourir l'arbre



R	N	U	O	B	J
10	01	00	111	1101	1100

Exercice

Soit X une source d'information codée sur l'ensemble de symboles $S = \{A, B, C, D, E, F, G\}$ dont les probabilités respectives sont:

A	B	C	D	E	F	G
0.350	0.300	0.200	0.100	0.040	0.005	0.005

1. Calculer l'auto information des symboles.
2. Calculer l'entropie de la source.
3. Déterminer le code de Huffman de la source
4. Calculer la longueur moyenne du code obtenu

Protection de l'information

- Eviter de transmettre l'information à des entités indésirables

  Forces armées en conflit
  Piraterie informatique
  Propriété intellectuelle

- Rendre robuste l'identification

  Banques et commerces en ligne
  Services officiels (passeports)

- Modifier le message transmis

  Sans changer l'information
  Seul l'émetteur et le récepteurs accèdent à l'information

Histoire

- XVI^{e} s av. J.-C. - Mésopotamie: tablettes en argile chiffrée

  Retrait de certaines consonnes
  Changement d'orthographe

- X^{e} s - IV^{e} s av. J.-C. - Antiquité Grecque

  Chiffrement par transposition à l'aide de Scytale

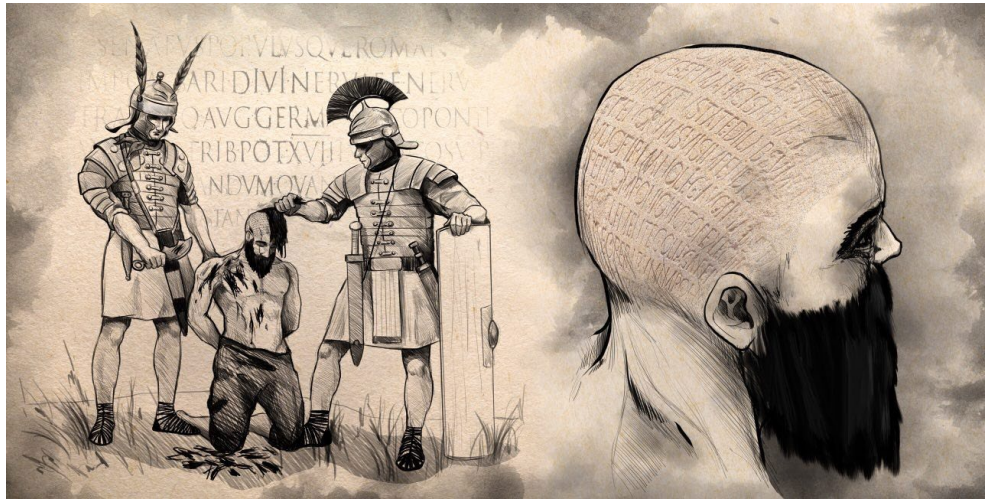
1. Choix de deux bâtons identiques
2. Découpage d'une bande de parchemin
3. Enroulement du parchemin autour du premier bâton
4. Ecriture du message
5. Déroulement du parchemin et envoi
6. Réception du parchemin et enroulement autour du second bâton
7. Lecture de message

Source: *The Codebreakers : A Comprehensive History of Secret Communication from Ancient Times to the Internet, Revised and Updated*
https://doc.lagout.org/security/Crypto/The%20CodeBreakers%20-%20Kahn%20David.pdf

Histoire

VIe s av. J.-C. - Mésopotamie

1. Raser la tête d'un esclave
2. Tatouer le message sur son crâne
3. Envoyer l'esclave à destination
4. Raser les cheveux de l'esclave
5. Lire le message



Ve s - IVe s av. J.-C. - Hebreux

Méthode de l'Atbash pour chiffrer les textes religieux

Atbash = **A**leph, **T**av, **B**eth, **Sh**in (premières et dernières lettres de l'alphabet)

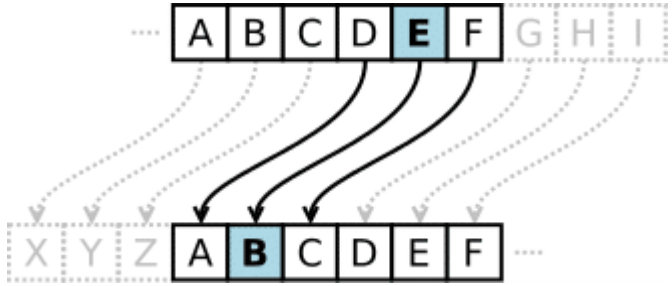
Consiste à remplacer chaque lettre du message par une autre selon une correspondance précise.

Histoire

Ie s av. J.-C. - Epoque romaine

Chiffrement de César: décalage les lettres de l'alphabet d'un nombre n. Première notion de clé.

1. Rédiger un texte : « décaler les lettres de l'alphabet ».
2. Choisir une clé: $n = 3$
3. chifffrer le texte : « ghfdohu ohv ohwwuhv gh o'doskdehw ».



Ve s - IVe s av. J.-C. - Hebreux

Méthode de l'Atbash pour chiffrer les textes religieux

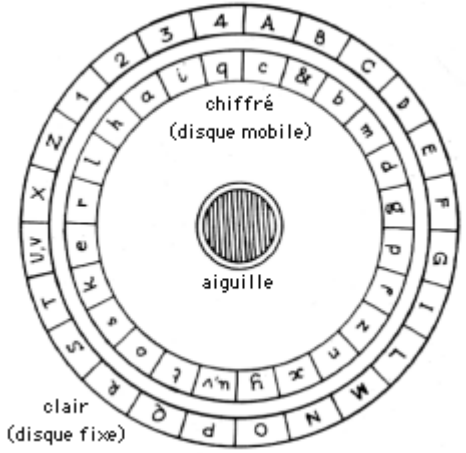
Atbash = **A**leph, **T**av, **B**eth, **Sh**in (premières et dernières lettres de l'alphabet)

Consiste à remplacer chaque lettre du message par une autre selon une correspondance précise.

Histoire

XV^{e} s - Moyen âge

Chiffrement par substitution polyalphabétique appliqué à l'aide d'un disque à chiffrer, ou cadrant chiffrant (*Leon Battista Alberti*).



Préparation:

1. Alphabet d'origine de 20 lettres (pas de H, J, K, U, W et Y).
2. Alphabet de codage de 23 lettres (pas de J, U, W).
3. **Disque externe**: 24 segments (alphabet d'origine + {1, 2, 3, 4}).
4. **Disque interne**: 24 segments (alphabet de codage dans un ordre arbitraire + {&}).

Codage:

1. Retirer les espaces du texte
2. Retirer ou substituer les lettres manquantes dans l'alphabet d'origine: J = I, K = Q, U, W = V, Y = I et H disparait.
3. Décalage initial du disque interne de n segments vers la droite.
4. Chiffrement du texte original en utilisant la correspondance entre le disque externe et le disque interne. Tous les 4 symboles encodés, déplacer le disque interne dans le sens des aiguilles d'une montre de 1 symbole.

Histoire

Exercice: Codage d'Alberti

Soit l'alphabet original d'Alberti: {A, B, C, D, E, F, G, I, L, M, N, O, P, Q, R, S, T, V, X, Z} et soit l'alphabet de codage {u, s, q, o, m, k, h, f, d, b, a, c, e, g, i, l, n, p, r, t, x, z, &, y}

Utiliser le code d'Alberti pour coder le message « Bonjour les L3 » en utilisant un décalage initial de 1.

Origine	A	B	C	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	V	X	Z	1	2	3	4
Codage	u	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y

Histoire

Exercice: Codage d'Alberti

Message d'origine: « Bonjour les L3 »

1. Retirer les espaces du texte: « BonjourlesL3 »

2. Retirer ou substituer les lettres manquantes dans l'alphabet d'origine: J = I, K = Q, U, W = V, Y = I et H disparait: « BoniovrlesL3 »

3. Appliquer le décalage initial:

Origine	A	B	C	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	V	X	Z	1	2	3	4
Codage	u	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y
Codage + 1	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y	u

Histoire

Exercice: Codage d'Alberti

Message d'origine: « Bonjour les L3 »
Message préparé: « BoniovrlesL3 »

Disques:

Origine	A	B	C	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	V	X	Z	1	2	3	4
Codage	u	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y
Codage + 1	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y	u

4. Chiffrement du texte original en utilisant la correspondance entre le disque externe et le disque interne.
Tous les 4 symboles encodés, déplacer le disque interne dans le sels des aiguilles d'une montre de 1 symbole.

Codage des 4 premiers symboles:

B	o	n	i	o	v	r	l	e	s	L	3
q	e	c	d								

Histoire

Exercice: Codage d'Alberti

Message d'origine: « Bonjour les L3 »
Message préparé: « BoniovrlesL3 »

Décalage du disque interne d'un symbole vers la droite

Origine	A	B	C	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	V	X	Z	1	2	3	4
Codage	u	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y
Codage + 1	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y	u
Codage + 2	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y	u	s

Codage des 4 symboles suivants:

B	o	n	i	o	v	r	l	e	s	L	3
q	e	c	d	g	t	n	a				

Histoire

Exercice: Codage d'Alberti

Message d'origine: « Bonjour les L3 »
Message préparé: « BoniovrlesL3 »

Décalage du disque interne d'un symbole vers la droite

Origine	A	B	C	D	E	F	G	I	L	M	N	O	P	Q	R	S	T	V	X	Z	1	2	3	4
Codage	u	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y
Codage + 1	s	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y	u
Codage + 2	q	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y	u	s
Codage + 3	o	m	k	h	f	d	b	a	c	e	g	i	l	n	p	r	t	x	z	&	y	u	s	o

Codage des 4 symboles suivants:

B	o	n	i	o	v	r	l	e	s	L	3
q	e	c	d	g	t	n	a	f	r	c	s

Voir: https://www.dcode.fr/chiffre-alberti