# DÉCRYPTAGE DU CHIFFREMENT DE VIGENÈRE

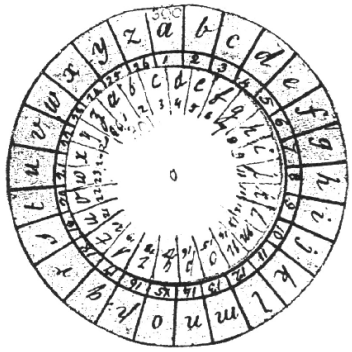
I. Lompo, L. de Maillard, F. Montiton, J-F. Morreeuw

Maîtrise Informatique — Université Bordeaux I

20 mai 1997

# De l'apparition du chiffrement

## Léone Battista Alberti (1466)



*Disque de chiffrement*

## Johannes Trithemius (1508)

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

*Tableau de chiffrement*

# Blaise de Vigenère (1586)

## Principe du chiffrement de Cæsar

- chiffrement par permutation de l'alphabet
- les lettres codent les 26 décalages possibles

```
A pas de décalage
B permutation d'une lettre de l'alphabet
    A devient B, B devient C ...
 ...
Z permutation de 25 lettres de l'alphabet
    A devient Z, B devient A ...
```

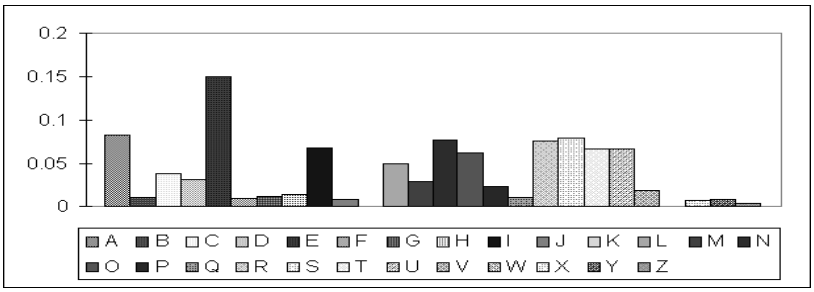
## Adaptation du chiffrement de Vigenère

- un mot de $n$ lettres code $n$ décalages
- ce mot est utilisé cycliquement pour servir de clé
- chaque lettre est décalée en fonction de la lettre de la clé qui lui correspond

# Défaut du chiffrement de Cæsar

## Faiblesse du chiffrement de Cæsar

- calcul des fréquences des lettres
- comparaison avec une langue connue
- déduction du décalage utilisé




*Fréquences des lettres en français*

## Solidité du chiffrement de Vigenère

- plusieurs alphabets décalés utilisés
- pas de déduction directe possible

## Mise en forme mathématique

Gauss (1801)
Introduit la notion de modulo

Charles Babbage (1846)

```
Cypher=Key+Translation-1
Translation=Cypher-Key+1
```

## Méthode dite de Kerckhoff (1883)

- clé utilisée de manière cyclique
- une position du cycle correspond à un décalage
- analyse fréquentielle possible pour une position
- méthode fastidieuse pour un calcul à la main si la longueur de la clé n'est pas connue

# Méthode dite de Kasiski (1863)

- plusieurs occurrences d'un mot peuvent apparaître dans le texte chiffré
- ces occurrences peuvent provenir d'un même mot
- la longueur de la clé divise les distances entre les occurrences dans le texte chiffré

```
wakeupalicedearsaidh    hegmmaehquphaijdeelz
ersisterwhywhatalong    pvoqkeinezjadillpkvy
sleepyouvehadohiveha    dpamhjsqdwsezwztzaps
dsuchacuriousdREAmsA    owqkzlgqzazyolJPEiaS
IDaliceandshetolDHER    THwtaniwvvdlabgwHDMJ
SISTERaswellasshecou    DMOBWCeoewwpwaksiywm
ldrememberthemallthe    whnmepqxmjelauswpppw
sestrangeadventureso    diobjlrcmsozavlfvaag
fhersthatyouhavejust    qlazkelwbqzydinpnqal
beenJPEdingaboutandw    miavJPEzqfrexwmeejlo
henSHEhadfiniSHEDHER    sijAZPlwlxtreAZPHDMJ
SISTERkissEDHerandsA    DMOBWCoeakPHDmjlrzaS
IDitwasacuriousdream    THebolwwkmcmkckovaie
dearcertainlybutnowr    oiwzupvpiaypujmerkej
unintoyourteaitsgett    frevlzckcjeiwqldkabl
inglate                 trctsei
```

- `STH` 160
- `HDMJDMOBWC` 120
- `JPE` 110
- `PHD` 15
- `AZP` 10

# Test de Friedman (1920)

Méthodes utilisées classiquement

Généralisation de la méthode de Kerckhoff

- étudier les lettres les plus fréquentes
- comparer globalement les répartitions

Simplifications mathématiques

- indice de coïncidence

$$I_c(x) \approx \sum_{i=0}^{25} p_i^2$$

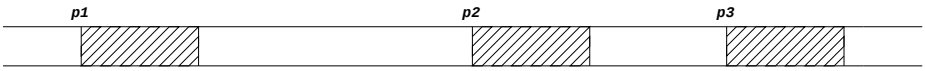
$I_c(aléatoire) = .03846$
$I_c(suédois) = .07063$
$I_c(espagnol) = .07393$
$I_c(italien) = .07565$

- indice de coïncidence mutuel

$$MI_c(x,y) \approx \sum_{i=0}^{25} p_i(x) . p_{(i+décalage)(mod\,26)}(y)$$

# Test de Kasiski

## Recherche des distances entre les occurrences




- limitation de l'étendue des recherches
- prise en compte des occurrences consécutives

## Déduction de la longueur de la clé

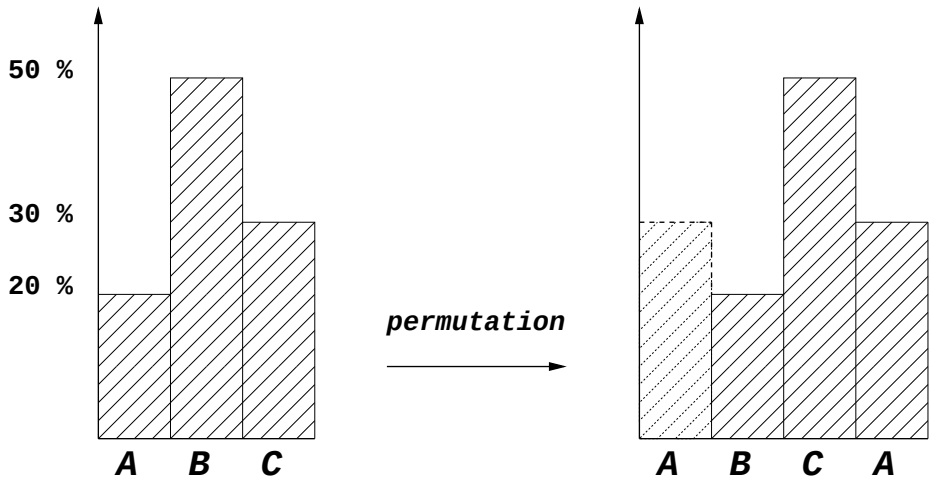
| | |
|---|---|
| `structure initiale` | `(160,1) (120,1) (110,1) (15,1) (10,1)` |
| $pgcd(160,120) = 40$ | `(160,1) (120,1) (110,1) (40,1) (15,1) (10,1)` |
| $pgcd(160,110) = 10$ | `(160,1) (120,1) (110,1) (40,1) (15,1) (10,2)` |
| $pgcd(160,40) = 40$ | `(160,1) (120,1) (110,1) (40,2) (15,1) (10,2)` |
| $pgcd(160,15) = 5$ | `(160,1) (120,1) (110,1) (40,2) (15,1) (10,2) (5,1)` |
| $pgcd(160,10) = 10$ | `(160,1) (120,1) (110,1) (40,2) (15,1) (10,3) (5,1)` |
| $pgcd(160,5) = 5$ | `(160,1) (120,1) (110,1) (40,2) (15,1) (10,3) (5,2)` |
| $pgcd(120,110) = 10$ | `(160,1) (120,1) (110,1) (40,2) (15,1) (10,3) (5,2)` |
| $pgcd(120,40) = 40$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,3) (5,2)` |
| $pgcd(120,10) = 10$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,4) (5,2)` |
| $pgcd(120,5) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,4) (5,3)` |
| $pgcd(110,40) = 10$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,5) (5,3)` |
| $pgcd(110,15) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,5) (5,4)` |
| $pgcd(110,10) = 10$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,6) (5,4)` |
| $pgcd(110,5) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,6) (5,5)` |
| $pgcd(40,15) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,6) (5,8)` |
| $pgcd(40,10) = 10$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,9) (5,8)` |
| $pgcd(40,5) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,9) (5,11)` |
| $pgcd(15,10) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,9) (5,12)` |
| $pgcd(15,5) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,9) (5,13)` |
| $pgcd(10,5) = 5$ | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,9) (5,22)` |
| `structure finale` | `(160,1) (120,1) (110,1) (40,3) (15,1) (10,9) (5,22)` |

# Méthode de Kerckhoff (1/4)

Recherche des symétries de position

- la longueur de la clé est supposée connue
- chaque lettre de la clé détermine une permutation
- connaître les décalages entre les alphabets permet de déterminer les décalages relatifs entre les lettre de la clé

Mise en équations de l'ensemble des possibilités




*Permutation d'un alphabet de trois lettres*

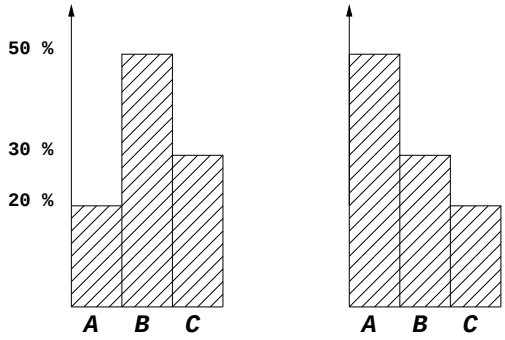
# Méthode de Kerckhoff (2/4)

Choix du coefficient associé à chaque équation

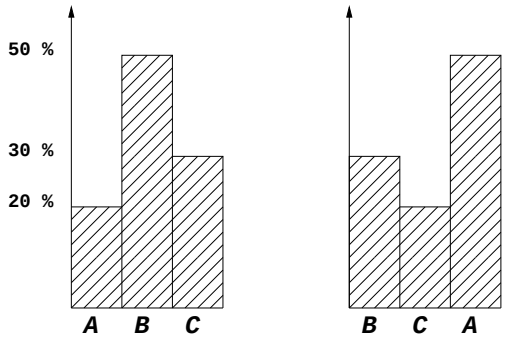
L'indice de coïncidence mutuel calcule la probabilité pour que deux répartitions fréquentielles de deux alphabets correspondent

- supposer une permutation donnée entre 2 alphabets
- rétablir la position des lettres par permutation inverse
- calculer la probabilité pour que les 2 alphabets correspondent
- la probabilité permet de juger l'hypothèse

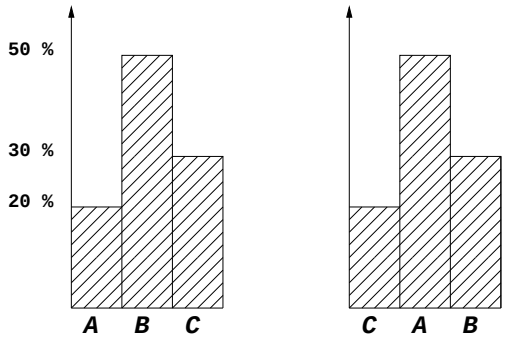
# Méthode de Kerckhoff (3/4)




.2*.5+.5*.3+.3*.2=.31




.2*.3+.5*.2+.3*.5=.31




.2*.2+.5*.5+.3*.3=.38

La permutation la plus probable est :

- A permuté en C
- B permuté en A
- C permuté en B

Si la lettre pour l'alphabet 1 est A, celle du 2 est C

# Méthode de Kerckhoff (4/4)

## Traduction en un système d'équations

- $d_{12} = 0$ associé au coefficient $MI^0_{c12}$
- $d_{12} = 1$ associé au coefficient $MI^1_{c12}$
- $d_{12} = 2$ associé au coefficient $MI^2_{c12}$
- $d_{23} = 0$ associé au coefficient $MI^0_{c23}$, $d_{23} = 1 \ldots$
- $d_{13} = d_{12} + d_{23} = 0$ associé au coefficient $MI^0_{c13}$, $d_{13} = 1 \ldots$

## Recherche des meilleures solutions pour le système

- hypothèse pour le décalage entre les lettres de la clé
- sélection des équations correspondant à l'hypothèse
- traduction de cette hypothèse en un coefficient

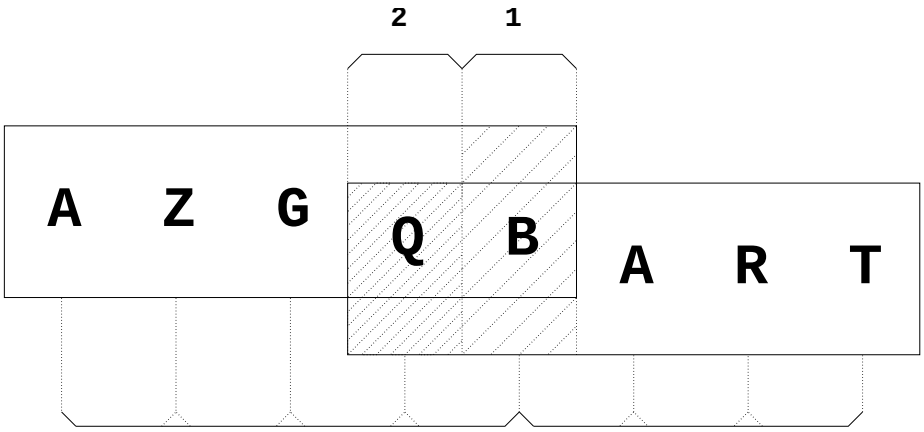
Par exemple, si $d_{12} = 2$, $d_{23} = 1$, la sélection est :

- $d_{12} = 2$ associé au coefficient $MI^2_{c12}$
- $d_{23} = 1$ associé au coefficient $MI^1_{c23}$
- $d_{13} = d_{12} + d_{23} = 0$ associé au coefficient $MI^0_{c13}$

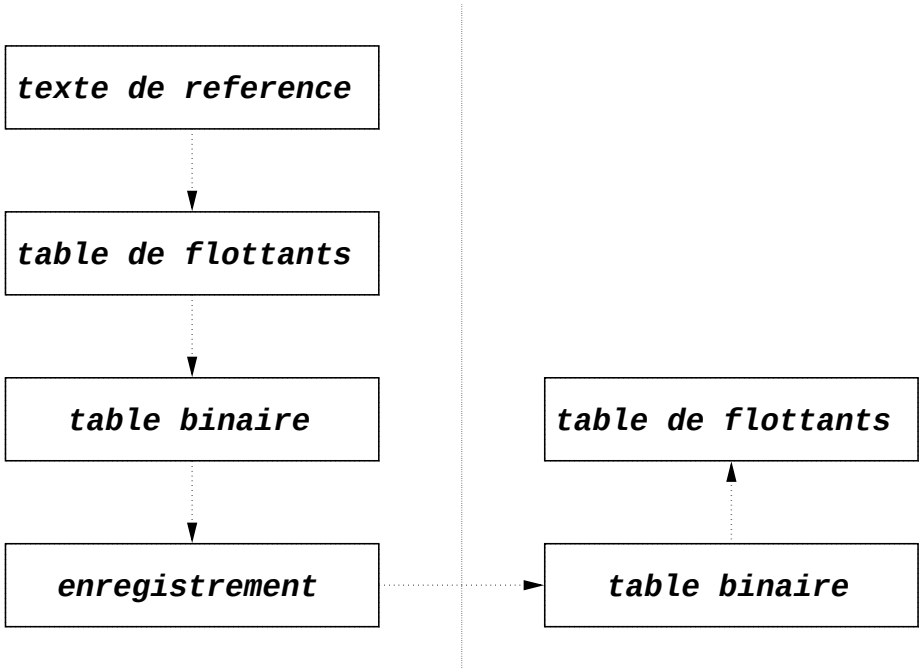
Et le coefficient de l'hypothèse est $MI^2_{c12} * MI^1_{c23} * MI^0_{c13}$

# Implantation de la résolution du système

```
result_p
_analyse(char *data, long length, short cycle, short start,
         short distance, int depth)
{
  result_p result;   /* structure générale de résultats */

  if(distance <= KH_WIDTH_MAX)
    {
      ...   /* mise en place des données de calcul */

      distance = KH_WIDTH_MAX;

      ...   /* recherche des meilleures solutions  */
    }
  else
    {
      long width1 = distance/2;
      long width2 = distance-width1;
      long distance1 = width1+1;   /* lettre de jointure     (1) */
      long distance2 = width2+1;   /* lettre de vérification (2) */
      result_p result1 = _analyse(data, length, cycle, start, distance1, depth);
      result_p result2 = _analyse(data, length, cycle,
                                  start+width1-1, distance2, depth);

      result = result_join(result1, width1, 0, width1-1,
                           result2, width2-1, 1, 0);

      result_destroy(result1);
      result_destroy(result2);
    }
  return result;
}
```



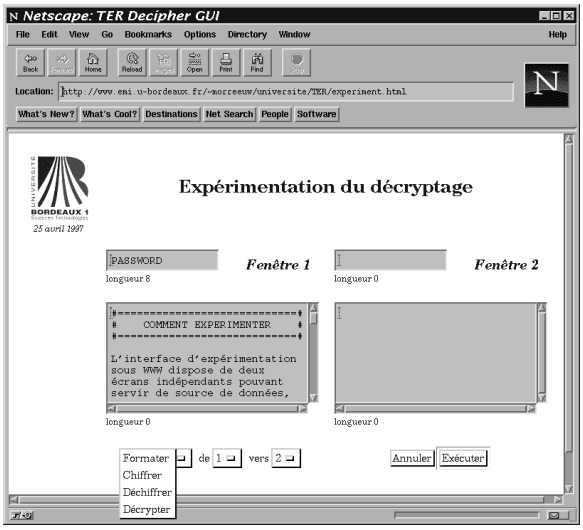
# Portabilité des données



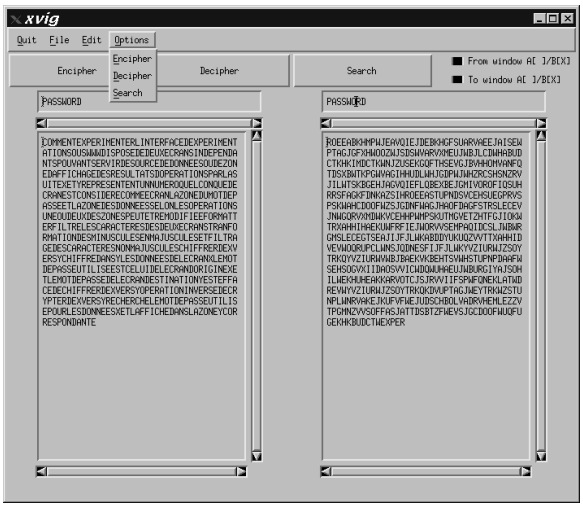

Utilisation de la norme IEEE 754 simple précision

# Interfaces d'expérimentation

## Interface HTML




## Interface Xwindow

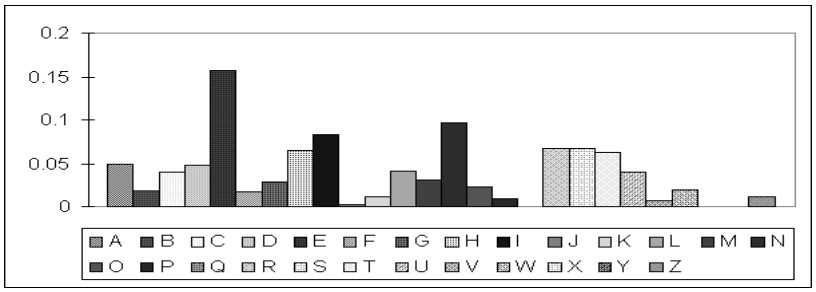


# Informations sur les langues

## Indices de coïncidence

- $I_c\,(esperanto) = .06883$
- $I_c\,(suédois) = .07063$
- $I_c\,(allemand) = .07187$
- $I_c\,(norvégien) = .07350$
- $I_c\,(espagnol) = .07393$
- $I_c\,(francais) = .07405$
- $I_c\,(italien) = .07565$

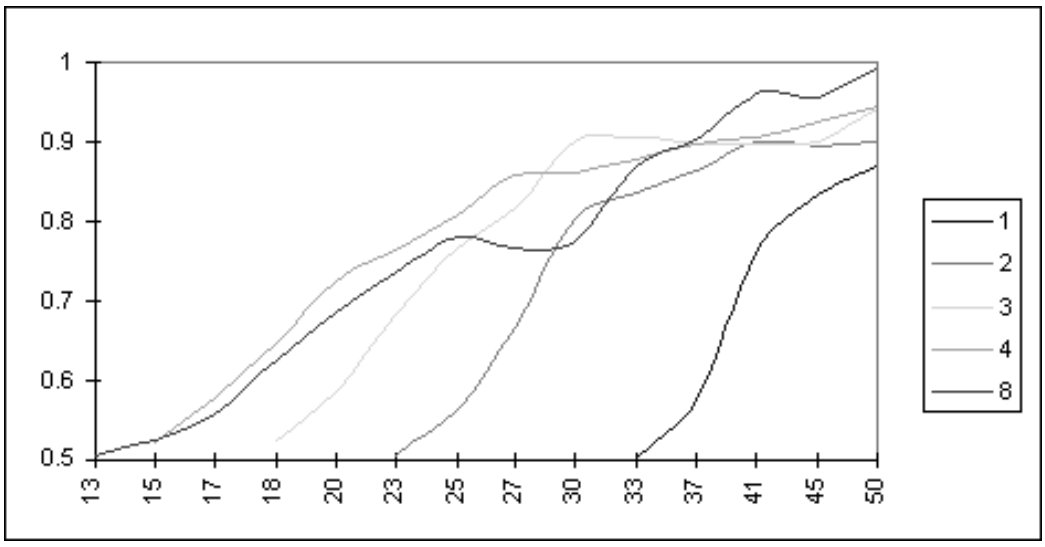
## Répartitions fréquentielles




*Fréquences de l'allemand*

# Performances des algorithmes

## Possibilités de décryptage




*Probabilités des succès en fonction du rapport $|texte|/|clé|$*

## Complexité des algorithmes

$$kh\_compute \leq |clé|\, log_2|clé| * depth\,(depth + N) + \frac{|clé|}{N}\,|texte| + depth\,|texte|\,N.K^N$$

$$search\_key\_length \leq |texte|\, buffer.log_2\,(buffer) + buffer * depth$$

## Rappel du sujet

- Statistiques des fréquences de lettres et de bi-lettres en français à partir de textes existant sur Internet
- Module de chiffrage/déchiffrage pour le Vigenère
- Implantation des tests de Kasiski et Friedman
- Essais de décryptage partir de textes de longueurs variées
- Conclusions sur la solidité du cryptage

## Plan de l'exposé

**A - Présentation historique des méthodes**

**B - Adaptation algorithmique des méthodes**

- Recherche de la longueur de la clé (Kasiski)
- Recherche des décalages entre les lettres de la clé (Kerckhoff)

**C - Description de quelques points techniques**

- Calcul des solutions optimisant un système d'équations
- Portabilité de données précalculées

**D - Présentation des résultats obtenus**

- Informations calculées sur les langues
- Possibilités de décryptage offertes par les méthodes