

# 1. Fouille de textes



# Plan

- Traiter un texte
- Lancer une requête simple
- Requêtes et concordances
- Les expressions régulières
- Les symboles NooJ
- Analyses statistiques d'une concordance

# Interroger un texte: TEXT > Locate

The screenshot shows the NooJ software interface. The main window displays a text file named "\_La femme de trente ans.not" with a search count of 1 / 926 TUs. The dialog box "Locate a pattern in \_La femme de trente ans" is open, showing options for pattern types (string, PERL, NooJ regular expression, NooJ grammar), a search field containing "table", and options for syntactic analysis, index (shortest/longest/all matches), and limitation (only 100 matches, all matches, 1 example per match). A "Reset Concordance" checkbox is checked, and a NooJ logo is visible at the bottom right of the dialog.

File Edit Lab Project Windows Info

\_La femme de trente ans.not

- 1 + / 926 TUs

Show Text Annotation Structure

Au commencement du mois d'avril 1813, il y  
beaux jours où les Parisiens voient pour la pr  
sans nuages. Avant midi, un cabriolet à pomp  
rue de Rivoli par la rue Castiglione, et s'arrêta  
nouvellement ouverte au milieu de la terrasse  
homme en apparence soucieux et maladif; de  
jaune et le faisaient vieux avant le temps; il jet  
descendit pour prendre dans ses bras une jeu  
oisifs en promenade sur la terrasse. La petite  
quand elle fut debout sur le bord de la voiture  
posa sur le trottoir, sans avoir chiffonné la ga  
eu tant de soin. L'inconnu devait être le père  
familièrement le bras et l'entraîna brusquemen

Locate a pattern in \_La femme de trente ans

Pattern is:

- a string of characters:
- a PERL regular expression:
- a NooJ regular expression:
- a NooJ grammar:

table

Set

Syntactic Analysis

Index

- Shortest matches
- Longest matches
- All matches

Limitation

- Only: 100 matches
- All matches
- 1 example per match

Reset Concordance

N o o J

Cancel

# Interroger un texte: TEXT > Locate

The screenshot shows the NooJ software interface. The main window displays a text document titled "\_La femme de trente ans.not" with a search count of 1 / 926 TUs. The dialog box "Locate a pattern in \_La femme de trente ans" is open, showing the search configuration. The pattern is "table", and the search type is "a NooJ regular expression". The "Index" section is set to "Longest matches", and the "Limitation" section is set to "Only: 100 matches". A red circle highlights the NooJ logo at the bottom right of the dialog box.

File Edit Lab Project Windows Info

\_La femme de trente ans.not

- 1 + / 926 TUs

Show Text Annotation Structure

Au commencement du mois d'avril 1813, il y  
beaux jours où les Parisiens voient pour la pr  
sans nuages. Avant midi, un cabriolet à pomp  
rue de Rivoli par la rue Castiglione, et s'arrêta  
nouvellement ouverte au milieu de la terrasse  
homme en apparence soucieux et maladif; de  
jaune et le faisaient vieux avant le temps; il jet  
descendit pour prendre dans ses bras une jeu  
oisifs en promenade sur la terrasse. La petite  
quand elle fut debout sur le bord de la voiture  
posa sur le trottoir, sans avoir chiffonné la ga  
eu tant de soin. L'inconnu devait être le père  
familièrement le bras et l'entraîna brusquemen

Locate a pattern in \_La femme de trente ans

Pattern is:

a string of characters:

a PERL regular expression:

a NooJ regular expression:

table

a NooJ grammar:

Set

Syntactic Analysis

Index

Shortest matches

Longest matches

All matches

Limitation

Only: 100 matches

All matches

1 example per match

Reset Concordance

N o o J

Cancel

# > Concordance

Noo! - [Concordance for Text \_La femme de trente ans.not]

File Edit Lab Project Windows Info TEXT CONCORDANCE

Reset Display:   characters before, and  after. Display:  Matches  Outputs  
 word forms

Text	Before	Seq.	After
	puis elle prit une petite	table	, chercha du papier et se
	lentement ses lunettes sur la	table	, y remit aussitôt la lettre
	vous la lisiez... - Si à	table	un mets ne nous semble
	plaisanteries. En se mettant à	table	, les deux femmes regardèrent simultanément
	femmes se levaient-elles de	table	, que le valet de chambre
	'Aiglemont en posant sur une	table	la tasse dans laquelle il
	dîner se mettre à une	table	, y regarder les mets avec
	faiblesse ? Plus loin, devant une	table	ronde éclairée par des lampes
	méditative, un coude sur la	table	et la tête appuyée sur
	à l'avenir. Entre cette	table	et la marquise, une grande
	éclipsait un Drolling. Sur une	table	en laque de Chine se

Query 11/11

Cancel

# > Concordance

The screenshot shows the NooJ interface for a concordance search. The window title is "NooJ - [Concordance for Text \_La femme de trente ans.not]". The menu bar includes File, Edit, Lab, Project, Windows, Info, TEXT, and CONCORDANCE. The control panel features a "Reset" button, a "Display:" dropdown set to "word forms", radio buttons for "characters" and "word forms" (the latter is selected), the text "before, and" followed by a numeric input field containing "5" and the word "after.", a "display:" checkbox which is checked, and checkboxes for "Matches" and "Outputs". The main display area is a table with columns "Text", "Before", "Seq.", and "After". The search results are displayed in red text. The "Text" column contains the surrounding text, "Before" contains the word "table", "Seq." contains a comma, and "After" contains the following text. A red circle highlights the "Display:" dropdown and the "word forms" radio button. At the bottom, there is a "Query" field and a "Cancel" button.

Text	Before	Seq.	After
puis elle prit une petite	table	,	chercha du papier et se
lentement ses lunettes sur la	table	,	y remit aussitôt la lettre
vous la lisiez... - Si à	table	,	un mets ne nous semble
plaisanteries. En se mettant à	table	,	les deux femmes regardèrent simultanément
femmes se levaient-elles de	table	,	que le valet de chambre
'Aiglemont en posant sur une	table	,	la tasse dans laquelle il
dîner se mettre à une	table	,	y regarder les mets avec
faiblesse ? Plus loin, devant une	table	,	ronde éclairée par des lampes
méditative, un coude sur la	table	,	et la tête appuyée sur
à l'avenir. Entre cette	table	,	et la marquise, une grande
éclipsait un Drolling. Sur une	table	,	en laque de Chine se

Query 11/11

Cancel

# > Concordance

The screenshot shows the NooJ software interface for a concordance search. The window title is "NooJ - [Concordance for Text \_La femme de trente ans.not]". The menu bar includes "File", "Edit", "Lab", "Project", "Windows", "Info", "TEXT", and "CONCORDANCE". The toolbar contains a "Reset" button, a "Display:" dropdown set to "word forms", radio buttons for "characters" and "word forms" (selected), a "before, and" field set to "5", a "Display:" dropdown with "Matches" selected, and an "Outputs" checkbox. The main area is a table with columns "Text", "Before", "Seq.", and "After". The "Before" column is circled in red. The search results are displayed in red text. The bottom status bar shows "Query" and "11/11".

Text	Before	Seq.	After
	puis elle prit une petite	table	, chercha du papier et se
	lentement ses lunettes sur la	table	, y remit aussitôt la lettre
	vous la lisiez... - Si à	table	un mets ne nous semble
	plaisanteries. En se mettant à	table	, les deux femmes regardèrent simultanément
	femmes se levaient-elles de	table	, que le valet de chambre
	'Aiglemont en posant sur une	table	la tasse dans laquelle il
	dîner se mettre à une	table	, y regarder les mets avec
	faiblesse ? Plus loin, devant une	table	ronde éclairée par des lampes
	méditative, un coude sur la	table	et la tête appuyée sur
	à l'avenir. Entre cette	table	et la marquise, une grande
	éclipsait un Drolling. Sur une	table	en laque de Chine se

Query 11/11

Cancel

# > Concordance

NooJ - [Concordance for Text \_La femme de trente ans.not]

File Edit Lab Project Windows Info TEXT CONCORDANCE

Reset Display:  characters before, and  after. Display:  Matches  Outputs

word forms

Text	Before	Seq.	After
	puis elle prit une petite	table	, chercha du papier et se
	lentement ses lunettes sur la	table	, y remit aussitôt la lettre
	vous la lisiez... - Si à	table	un mets ne nous semble
	plaisanteries. En se mettant à	table	, les deux femmes regardèrent simultanément
	femmes se levaient-elles de	table	, que le valet de chambre
	'Aiglemont en posant sur une	table	la tasse dans laquelle il
	dîner se mettre à une	table	, y regarder les mets avec
	faiblesse ? Plus loin, devant une	table	ronde éclairée par des lampes
	méditative, un coude sur la	table	et la tête appuyée sur
	à l'avenir. Entre cette	table	et la marquise, une grande
	éclipsait un Drolling. Sur une	table	en laque de Chine se

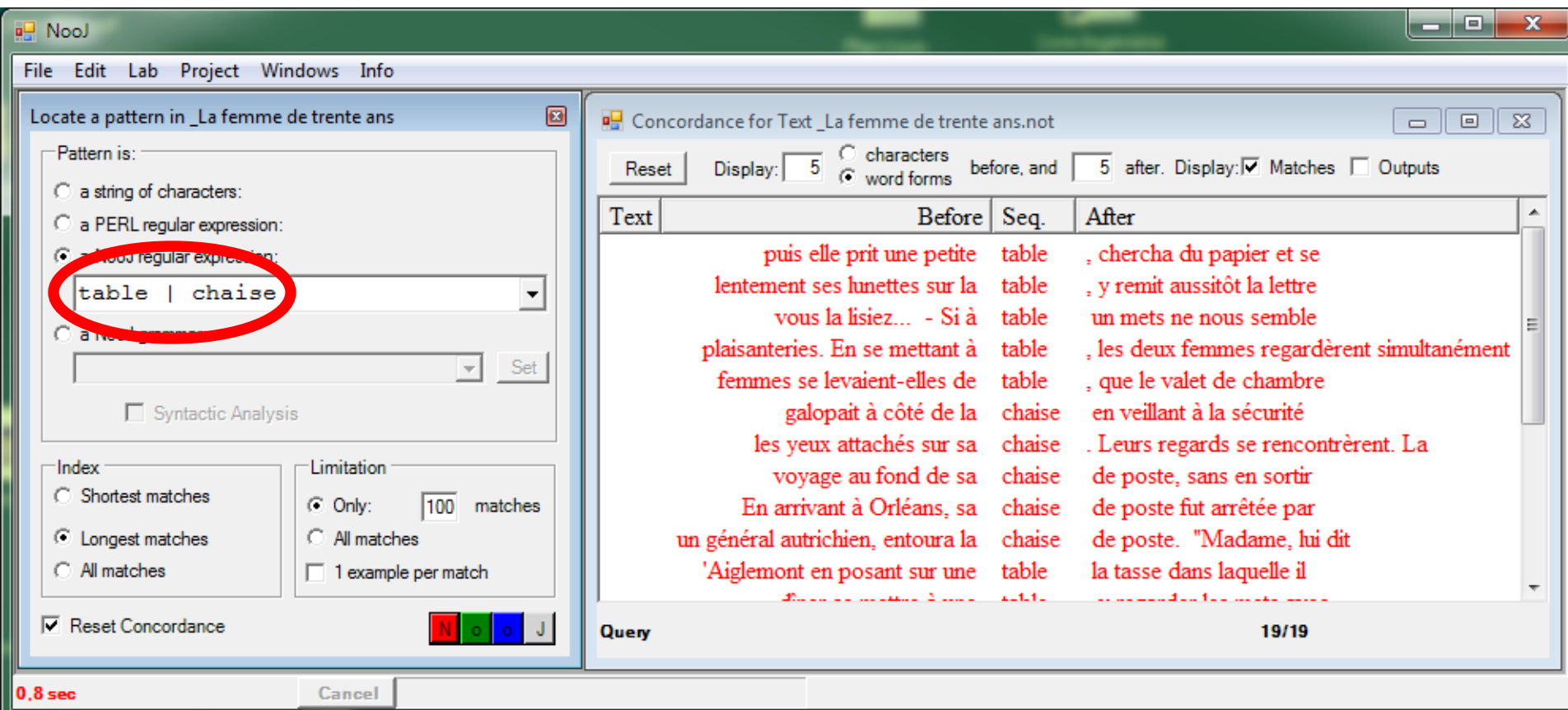
Query 11/11

Cancel



# Les expressions régulières

## 1. La disjonction, ex. **table | chaise**



The screenshot shows the NooJ interface with a search window titled "Locate a pattern in \_La femme de trente ans" and a concordance window titled "Concordance for Text \_La femme de trente ans.not".

**Search Configuration:**

- Pattern is:  a Perl regular expression: `table | chaise`
- Limitation:  Only: 100 matches
- Reset Concordance

**Concordance Results:**

Text	Before	Seq.	After
	puis elle prit une petite	table	, chercha du papier et se
	lentement ses lunettes sur la	table	, y remit aussitôt la lettre
	vous la lisiez... - Si à	table	un mets ne nous semble
	plaisanteries. En se mettant à	table	, les deux femmes regardèrent simultanément
	femmes se levaient-elles de	table	, que le valet de chambre
	galopait à côté de la	chaise	en veillant à la sécurité
	les yeux attachés sur sa	chaise	. Leurs regards se rencontrèrent. La
	voyage au fond de sa	chaise	de poste, sans en sortir
	En arrivant à Orléans, sa	chaise	de poste fut arrêtée par
	un général autrichien, entoura la	chaise	de poste. "Madame, lui dit
	'Aiglemont en posant sur une	table	la tasse dans laquelle il

Query 19/19

# Les expressions régulières

## 2. La concaténation, ex. **la maison**

The screenshot shows the NooJ software interface. On the left, a dialog box titled 'Locate a pattern in \_La femme de trente ans' is open. The 'Pattern is:' section has three radio buttons: 'a string of characters:', 'a PERL regular expression:', and 'a NooJ regular expression:'. The third option is selected, and the text 'la maison' is entered in the text box below it. The 'Index' section has three radio buttons: 'Shortest matches', 'Longest matches', and 'All matches'. The second option is selected. The 'Limitation' section has three radio buttons: 'Only:', 'All matches', and '1 example per match'. The first option is selected, and the value '100' is entered in the text box next to it. The 'Reset Concordance' checkbox is checked. On the right, a window titled 'Concordance for Text\_La femme de trente ans.not' is open. It has a 'Reset' button, a 'Display:' dropdown set to '5', and radio buttons for 'characters' and 'word forms'. The 'word forms' option is selected. There are also input boxes for 'before, and' and 'after', both set to '5'. The 'Display:' section has checkboxes for 'Matches' (checked) and 'Outputs'. The concordance table has four columns: 'Text', 'Before', 'Seq.', and 'After'. The text in the 'Text' column is highlighted in green. The 'Seq.' column contains the phrase 'la maison' repeated for each line. The 'Before' and 'After' columns contain the surrounding text from the original document. At the bottom of the window, the text 'Query' is on the left and '11/11' is on the right.

Text	Before	Seq.	After
et où les gens de enfants ou l'ami de train. Déjà le maître de la porte d'entrée de des arbres, mugissant autour de porte, réveillèrent les échos de bientôt a la porte de 'arrêtèrent à la porte de fit trembler les échos de salon, au dehors et dans timide précepteur, nouveau venu dans		la maison la maison la maison la maison la maison la maison la maison la maison la maison la maison la maison	plaçaient les bagages; mais ; mais ces personnes leur ga , de concert avec sa compa donnaient immédiatement s , ou s'engouffrant dans les . Ces coups prolongés eures où son fils le suivit . Après avoir échangé quelq . Puis il s'apaisa tout , n'était interrompu que par où il se trouvait, que

# Les expressions régulières

## 3. Priorité de la concaténation sur la disjonction, ex. **la (maison | chaise)**

The screenshot shows the NooJ software interface. The main window is titled "Concordance for Text \_La femme de trente ans.not". The search pattern is "la (maison | chaise)". The results are displayed in a table with columns "Text", "Before", "Seq.", and "After".

Text	Before	Seq.	After
	et où les gens de	la maison	plaçaient les bagages; mais alors
	qui galopait à côté de	la chaise	en veillant à la sécurité
	était un général autrichien, entoura	la chaise	de poste. "Madame, lui dit
	enfants ou l'ami de	la maison	; mais ces personnes leur gardent
	train. Déjà le maître de	la maison	, de concert avec sa compagne
	la porte d'entrée de	la maison	donnaient immédiatement sur le che
	des arbres, mugissant autour de	la maison	, ou s'engouffrant dans les
	porte, réveillèrent les échos de	la maison	. Ces coups prolongés eurent un
	bientôt a la porte de	la maison	où son fils le suivit
	'arrêtèrent à la porte de	la maison	. Après avoir échangé quelques par
	fit trembler les échos de	la maison	. Puis il s'anaisa tout

0.7 sec

La concaténation est similaire à la multiplication

La disjonction est similaire à l'addition

- La multiplication a priorité sur l'addition :

$$2 \times (3 + 4) = 2 \times 3 + 2 \times 4 = 14$$

$$\neq 2 \times 3 + 4 = 10$$

- La concaténation a priorité sur la disjonction :

$$\text{la (table | chaise) = la table | la chaise}$$

$$\neq \text{la table | chaise}$$

# Les expressions régulières

## 4. Factorisation, ex.

(ma | ta | sa | notre | votre | leur) (maison | chaise)

The screenshot shows the NooJ software interface. A dialog box titled "Locate a pattern in \_La femme de trente ans" is open. The "Pattern is:" section has three radio buttons: "a string of characters:", "a PERL regular expression:", and "a NooJ regular expression:". The "a NooJ regular expression:" option is selected, and the text "(ma | ta | sa | notre | votre | leur) (maison | chaise)" is entered in the text field. Below this, there is a "Set" button and a "Syntactic Analysis" checkbox. The "Index" section has three radio buttons: "Shortest matches", "Longest matches", and "All matches". The "Longest matches" option is selected. The "Limitation" section has three radio buttons: "Only:", "All matches", and "1 example per match". The "Only:" option is selected, and the number "100" is entered in the text field. The "Reset Concordance" checkbox is checked. At the bottom of the dialog, there are four colored buttons: a red button with 'N', a green button with 'o', a blue button with 'o', and a black button with 'J'. The background shows a window titled "trente ans.not" with a concordance table. The table has three columns: "Before", "Seq.", and "After". The "Seq." column contains the matches found by the regular expression. The "Before" and "After" columns contain the surrounding text. The table shows several rows of text with matches highlighted in blue. At the bottom of the window, there is a status bar showing "0,6 sec" and a "Cancel" button.

File Edit Lab Project Windows Info TEXT

Locate a pattern in \_La femme de trente ans

Pattern is:

a string of characters:

a PERL regular expression:

a NooJ regular expression:

(ma | ta | sa | notre | votre | leur) (maison | chaise)

a NooJ grammar:

Syntactic Analysis

Index

Shortest matches

Longest matches

All matches

Limitation

Only: 100 matches

All matches

1 example per match

Reset Concordance

0,6 sec

Cancel

trente ans.not

before, and 5 after. Display:  Matches  Outputs

Before	Seq.	After
hés sur	sa chaise	. Leurs regards se rencontrèrent. La
ond de	sa chaise	de poste, sans en sortir
rléans,	sa chaise	de poste fut arrêtée par
autres.	Sa maison	était, en toute chose, un
i rends	sa maison	agréable; je ferme les yeux
te dans	sa maison	et l'accompagna partout avec
là, que	sa chaise	vide parla pour lui, elle
ssé sur	sa chaise	. Une discussion avait fait traîner
iers de	sa maison	, et le fit entrer dans
sser de	sa maison	? " Ces idées fermentèrent soudain dans

10/10

# Les expressions régulières

On recherche **une petite table | une table**

5. Élément neutre :  $\varepsilon$  , "",  $\langle E \rangle$  :

**table  $\langle E \rangle$  = table ;  $\langle E \rangle$  table = table**

$\langle E \rangle$  est similaire à 1 pour la multiplication :

**17  $\times$  1 = 17 ; 1  $\times$  17 = 17**

# Les expressions régulières

## 5. Élément neutre <E>

**une petite table | une table**

**= une (petite table | table)**

# Les expressions régulières

## 5. Élément neutre $\langle E \rangle$

**une petite table | une table**  
**= une (petite table | table)**  
**= une (petite table |  $\langle E \rangle$  table)**



# Les expressions régulières

## 5. Élément neutre <E>

**une petite table | une table**  
**= une (petite table | table)**  
**= une (petite table | <E> table)**  
**= une ( (petite |<E>) table)**

# Les expressions régulières

## 5. Élément neutre $\langle E \rangle$

**une petite table | une table**  
**= une (petite table | table)**  
**= une (petite table |  $\langle E \rangle$  table)**  
**= une ( (petite |  $\langle E \rangle$ ) table)**

# Les expressions régulières

## 5. Élément neutre <E>

**une petite table | une table**  
**= une (petite table | table)**  
**= une (petite table | <E> table)**  
**= une ( (petite |<E>) table)**  
**= une (petite |<E>) table**

- Quand un élément XXX est facultatif, on écrira :

**(XXX | <E>)**

# Les expressions régulières

## 5. Élément neutre <E>

une (petite | <E>) table

The screenshot shows the NooJ software interface. On the left, a dialog box titled "Locate a pattern in \_La femme de trente ans" is open. It shows the pattern "une (petite | <E>) table" entered in the "a NooJ regular expression:" field. The "Reset Concordance" checkbox is checked. On the right, a window titled "Concordance for Text \_La femme de trente ans.not" displays the search results in a table format. The table has four columns: "Text", "Before", "Seq.", and "After". The results show the word "table" highlighted in green in the original image, with the surrounding text in the "Before" and "After" columns. The "Seq." column shows the sequence number of each match. At the bottom of the concordance window, it says "Query 5/5".

File Edit Lab Project Windows Info TEXT

Locate a pattern in \_La femme de trente ans

Pattern is:

- a string of characters:
- a PERL regular expression:
- a NooJ regular expression:  
une (petite | <E>) table
- a NooJ grammar:

Syntactic Analysis

Index

- Shortest matches
- Longest matches
- All matches

Limitation

- Only: 100 matches
- All matches
- 1 example per match

Reset Concordance

Concordance for Text \_La femme de trente ans.not

Reset Display: 5 characters before, and 5 after. Display:  Matches  Outputs

Text	Before	Seq.	After
	elle pensa; puis elle prit	une petite table	, chercha du papier et se
	d'Aiglemont en posant sur	une table	la tasse dans laquelle il
	appeler dîner se mettre à	une table	, y regarder les mets avec
	la faiblesse ? Plus loin, devant	une table	ronde éclairée par des lampes
	Dow éclipseait un Drolling. Sur	une table	en laque de Chine se

Query 5/5

0,6 sec Cancel

# Les expressions régulières

6. Opération de Kleene  $*$ , ex.

**une (<E> | très | très très | très très très | très  
très | très | très | ...) petite table**

# Les expressions régulières

6. Opération de Kleene  $*$ , ex.

une (<E> | très | très très | très très très | très  
très | très | très | ...) petite table =  
**une très\* petite table**

# Les expressions régulières

6. Opération de Kleene  $*$ , ex.

une (<E> | très | très très | très très très | très très très | très très très | ...) petite table =  
**une très\* petite table**

- L'opération de Kleene est indispensable en linguistique, par ex. :

**Sujet Verbe Complément\***  
**être adverbe\* participe passé**

# Les expressions régulières : récapitulation

cousin | cousine | cousins | cousines

(ma | ta | sa) (belle | <E>) (chaise | table)

(suis | es | est | sommes | êtes | sont) (en | dans | sous | sur)

(la | cette) très\* (belle | jolie) fleur

(la | cette) (très\* (belle | jolie) | <E>) fleur



# Les symboles de NooJ :

## 1. les symboles lexicaux

**maison | maisons**

**cousin | cousine | cousins | cousines**

# Les symboles de NooJ :

## 1. les symboles lexicaux

**maison | maisons = <maison>**

**cousin | cousine | cousins | cousines = <cousin>**

# Les symboles de NooJ :

## 1. les symboles lexicaux

**<être>** = être | été | étant |

suis | es | est | sommes | êtes | sont | étais | était |

étions | étiez | étaient |

fus | fut | fûmes | fûtes | furent |

serai | seras | sera | serons | serez | seront |

serais | serait | serions | seriez | seraient |

sois | soit | soyons | soyez | soient |

fusse | fusses | fût | fussions | fussiez | fussent

**ATTENTION :** <être> n'est qu'une abréviation. Si la forme *été* n'a pas été désambiguïsée, elle sera reconnue même dans « Je pars l'été prochain ». Pour désambiguïser les formes, on construira des grammaires qui décrivent leur contexte.

# Les symboles de NooJ :

## 1. les symboles lexicaux

**<manifester> =**

**manifester | manifeste | manifestons | ... |**  
**manifestation | manifestations |**  
**manifestant | manifestante | ... |**  
**remanifester | remanifestons | ... |**  
**manifestable | manifestables |**  
**manifestabilité | manifestabilités |**  
**...**

<manifester> réfère à l'entrée lexicale *manifester*. Si cette entrée est associée à ses formes conjuguées et dérivées, <manifester> les reconnaîtra. => C'est l'auteur du dictionnaire qui contrôle.

# Les symboles de NooJ :

## 1. les symboles lexicaux

**<manifester> =**

manifester | manifeste | manifestons | ... |  
manifestation | manifestations |  
manifestant | manifestante | ... |  
remanifester | remanifestons | ... |  
manifestable | manifestables |  
manifestabilité | manifestabilités |  
...

- L'utilisateur peut définir ses propres liens :

**<roi> =** roi | reine | tsar | king | président ...

# Les symboles de NooJ :

## 2. les symboles syntaxiques

**à | dans | de | en | par | vers | ...**

# Les symboles de NooJ :

## 2. les symboles syntaxiques

à | dans | de | en | par | vers | ... = <PREP>

# Les symboles de NooJ :

## 2. les symboles syntaxiques

à | dans | de | en | par | vers | ... = **<PREP>**

**<N>**, **<A>**, **<V>**, **<DET>**, **<PREP>**, **<CONJS>**, **<CONJC>**...

- L'utilisateur peut créer ses propres symboles, ex. :

**<MEDIC>**, **<POLITIQUE>**, **<SENTIMENT>**,  
**<VERBEPSYCHOLOGIQUE>** ...



# Les symboles de NooJ :

## 3. les propriétés

- Symbole lexical, ex. **<être+PR>**, **<cousin-s>**
- Symbole syntaxique, ex. **<N+Hum>**, **<V+INF>**, **<DET+f>**
- On peut combiner des propriétés :  
**<N+Hum-f+p>** ou **<V+F-3+s>**  
**<cousin-Hum+p>** ou **<être-INF+s>**
- L'utilisateur peut définir ses propres propriétés, ex. :  
**<MEDECINE+Maladie-Pneumo>**

# Les symboles de NooJ :

## 4. les négations

- <N+Hum> vs <N-Hum>
- <!N> <!N+Hum+Pol> <!N-Hum>
- +UNAMB  
sous-marin, N+UNAMB
- +EXCLUDE  
<N> (<A> | est/+EXCLUDE)  
(<N> | la/+EXCLUDE) <N>

# Les symboles de NooJ :

## Récapitulation

<être> <PREP> <N>

<être> <ADV>\* <PREP> <N>

<carte> (<bancaire> | de crédit | de débit |  
<bleu> | ...)

<DET> ( (très | <E>) <A> | <E>) <N> <A>\*

# Les symboles de NooJ :

## Récapitulation

**maison** → *maison Maison MAISON*

**Maison** → *Maison MAISON*

**MAISON** → *MAISON*

**mAisOn** → *mAisOn*

**"Maison"** → *Maison*

**"maison"** → *maison*

**<maison>** → *maison, Maison MAISON, maisons, Maisons  
MAISONS*

**<maisons>** → ?

**<Maison>** → ?

# Chercher un concept

- Champ lexical du concept *mort* :

# Chercher un concept

- Champ lexical du concept *mort* :

<assassiner> | <mourir> | <périr> | <tuer> |  
<veuf> | funérailles | <cadavre> | <meurtre> |  
<fatal> | <mortel> |  
fatalement | mortellement

# Chercher un concept

The screenshot shows the NooJ software interface. The main window is titled "NooJ" and has a menu bar with "File", "Edit", "Lab", "Project", "Windows", "Info", "TEXT", and "CONCORDANCE". Below the menu bar, there is a search bar with the text "Locate a pattern in \_La femme de trente ans". The search pattern is defined as "a NooJ regular expression" and is set to "décès | <décéder> | deuil | <fatal> | <mort> | <mortel> | <mourir> | <périr> | funérailles".

Below the search bar, there is a window titled "Concordance for Text \_La femme de trente ans.not". This window has a "Reset" button, a "Display" field set to "10", and radio buttons for "characters" and "word forms". It also has a "before, and" field set to "10" and an "after" field set to "10". There are checkboxes for "Display: Matches" (checked) and "Outputs".

The concordance results are displayed in a table with three columns: "Text", "Before", and "After". The text is highlighted in red. The "Before" column shows the text "treizième de l'année 1813 . Le surlendemain, Napoléon partait pour cette D'ici je puis encore apercevoir l'Empereur ; s'il Wagram au milieu du feu, à la Moscowa parmi les allez donc encore m'affliger en me parlant de votre te voir fidèle à ton vieux père jusqu'à sa un médaillon suspendu à son cou par une chaîne de sa tristesse dans le chagrin que lui avait causé la mort de son père, de qui elle portait encore le 'est mis à étudier sa maladie, que l'on croyait ne savent aimer; ils croient que d'aller à la pour nous. Autrefois, l'on savait aussi bien aimer que à souhaiter un divorce, si toutefois vous n'étiez pas elle perdit la comtesse de Listomère-Landon. La vieille dame la femme et du mari plus parfait, cette personne était à son maître imbécile, parer de fleurs une maison de attendit avec résignation la fin de ses peines en espérant". The "After" column shows the text "fatale campagne pendant laquelle il allait perdre successivement Bessières et Duroc périssait pendant la campagne, je ne l'aurais jamais vu." Le , il est toujours tranquille comme Baptiste lui !" Cette réponse à , J'étais si gaie ! Voulez-vous bien chasser vos vilaines , j'espérais te conserver près de moi heureuse et brillante , et le portrait de son père lui apparut soudain. A de son père, de qui elle portait encore le deuil , Au bout de huit jours, la douairière admira la douceur . Insensiblement, il a pris goût à l'anatomie, à la le lendemain les dispense d'avoir, la veille, des égards à propos. Ma nièce, je vous le formerai. Je mettrai avant d'en venir au désespoir." Julie écoutait sa tante de joie et d'une goutte remontée au coeur, en . Julie sentit toute l'étendue de cette perte. Il n , et afficher le bonheur sur un visage pâli par de ieune. Depuis son retour de Touraine, sa santé s'était".

At the bottom of the window, there is a "Query" field and a "Cancel" button. The status bar at the bottom left shows "5,8 sec" and the status bar at the bottom right shows "136/136".

# Analyse statistique, score standard

- Quantifier l'importance d'une information ?



# Analyse statistique, score standard

- Quantifier **l'importance** d'une information ?
- Pas par la fréquence, ex. **le/la** est très fréquent.

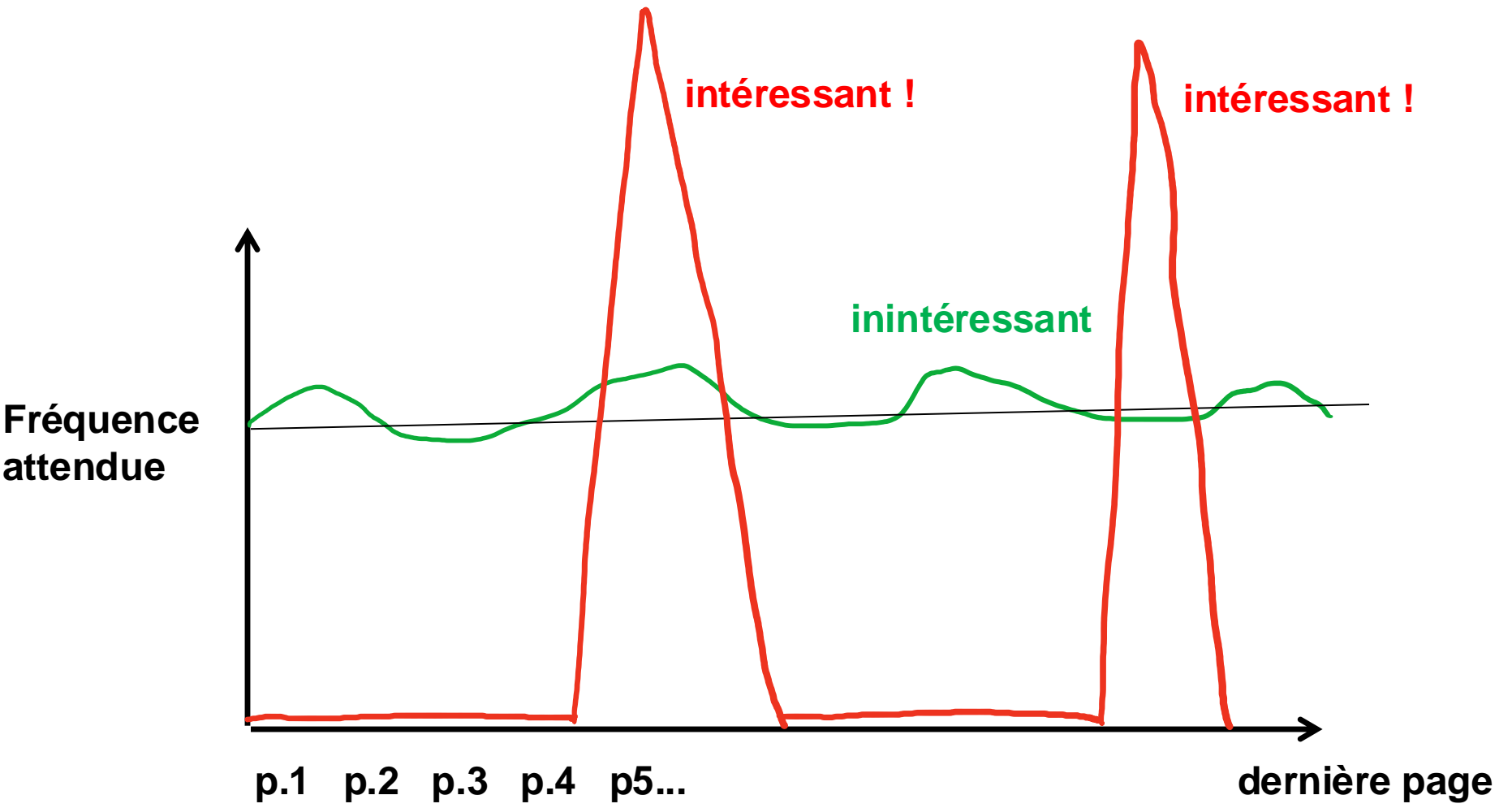
# Analyse statistique, score standard

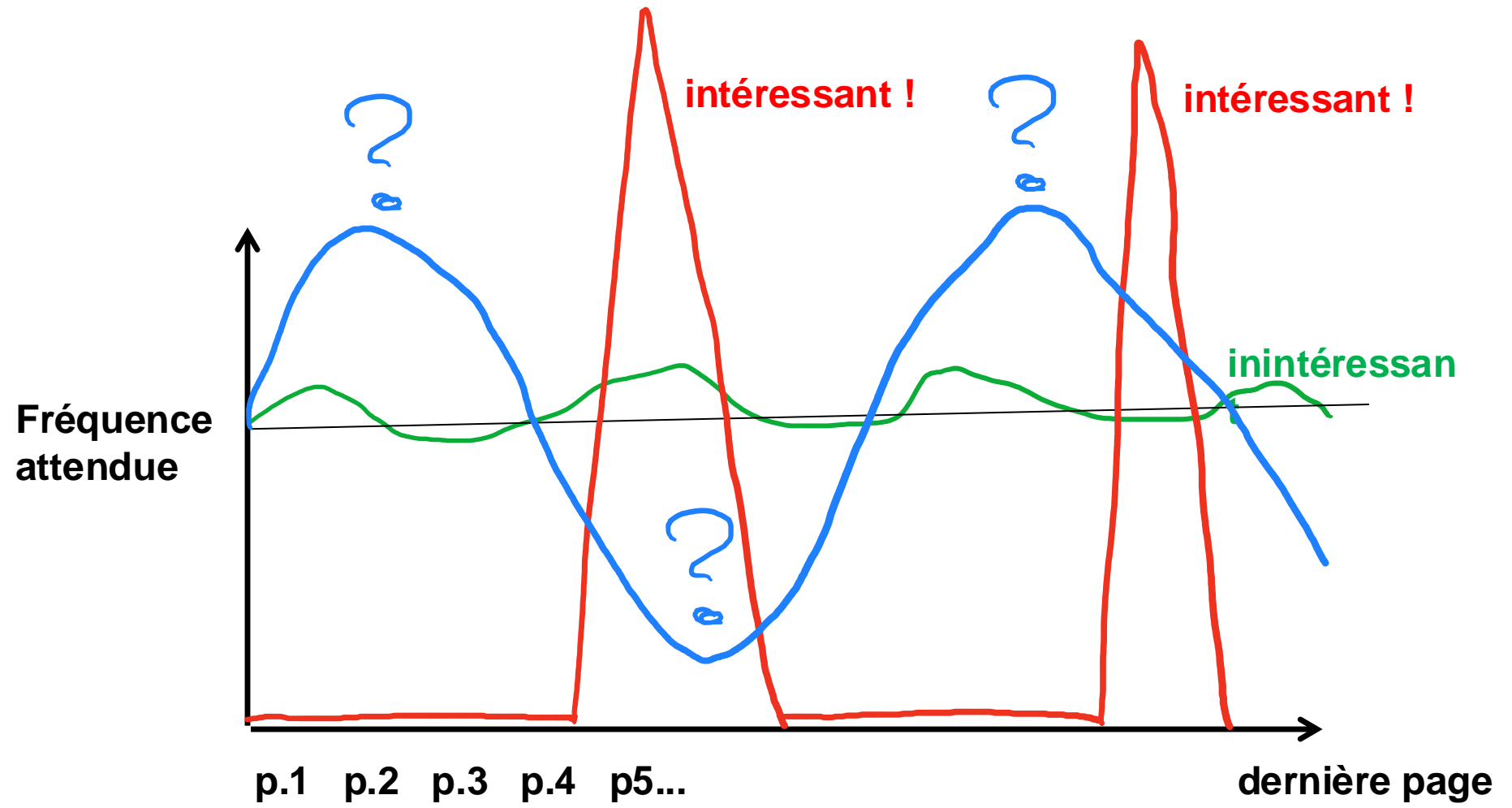
- Quantifier **l'importance** d'une information ?
- Pas par la fréquence, ex. **le/la** est très fréquent.
- Une seule occurrence du terme ***soucoupe volante*** dans un texte du XIXe siècle serait extrêmement intéressant !

# Analyse statistique, score standard

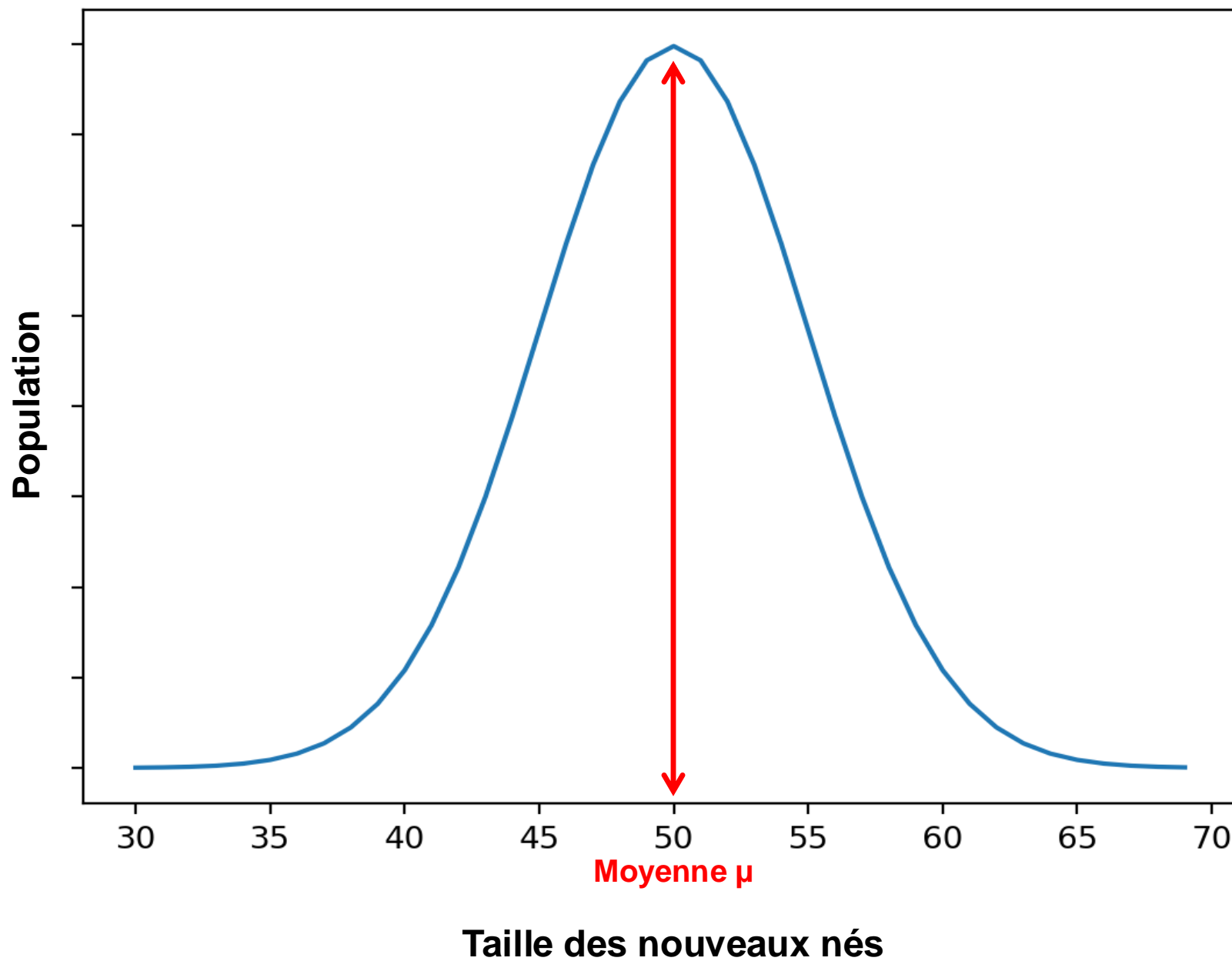
- Quantifier **l'importance** d'une information ?
- Pas par la fréquence, ex. *le/la* est très fréquent.
- Une seule occurrence du terme *soucoupe volante* dans un texte du XIXe siècle serait extrêmement intéressant !
- On va donc chercher des **accidents de fréquence**, i.e. des différences anormales entre la fréquence attendue et la fréquence absolue



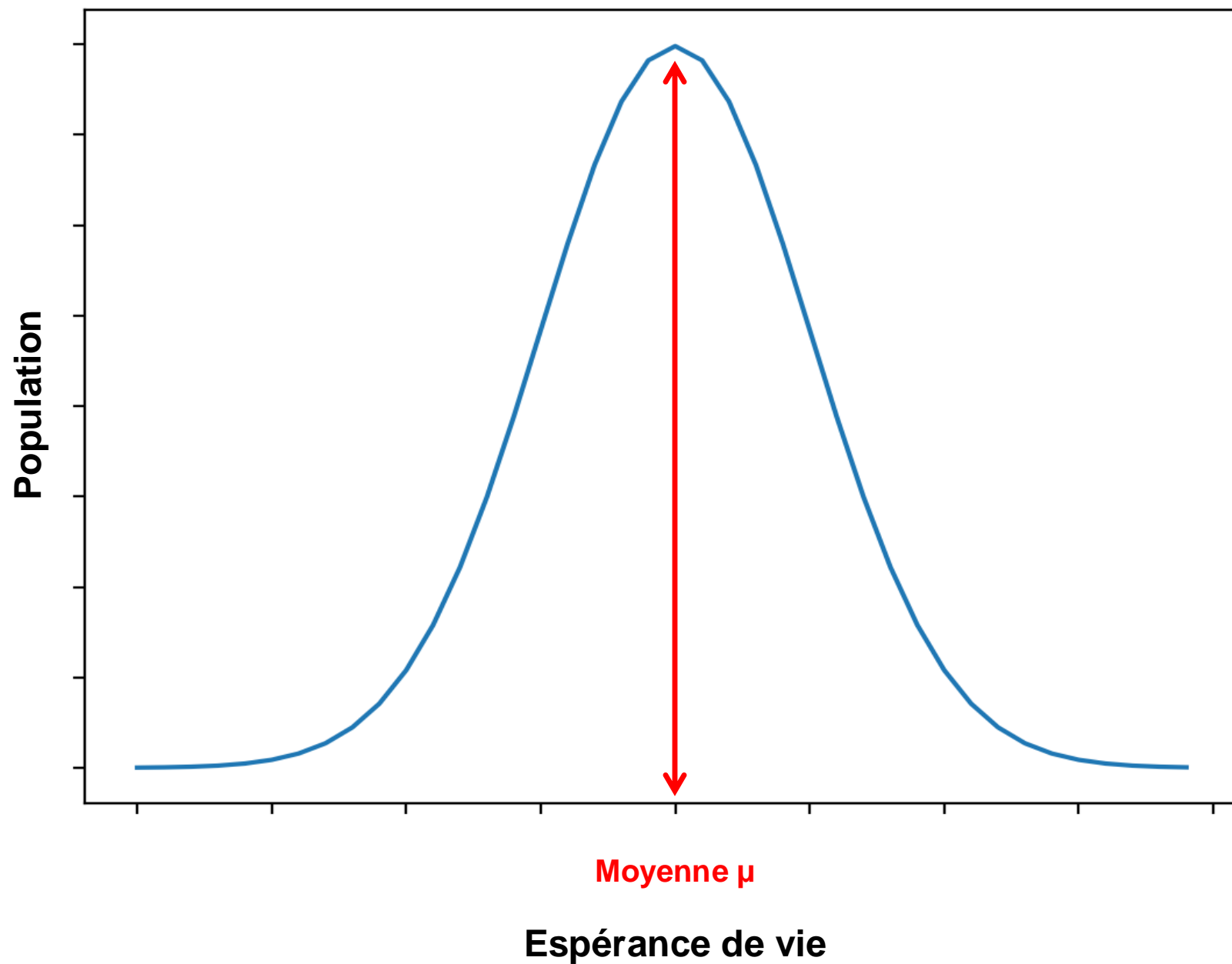




# La distribution normale de Gauss

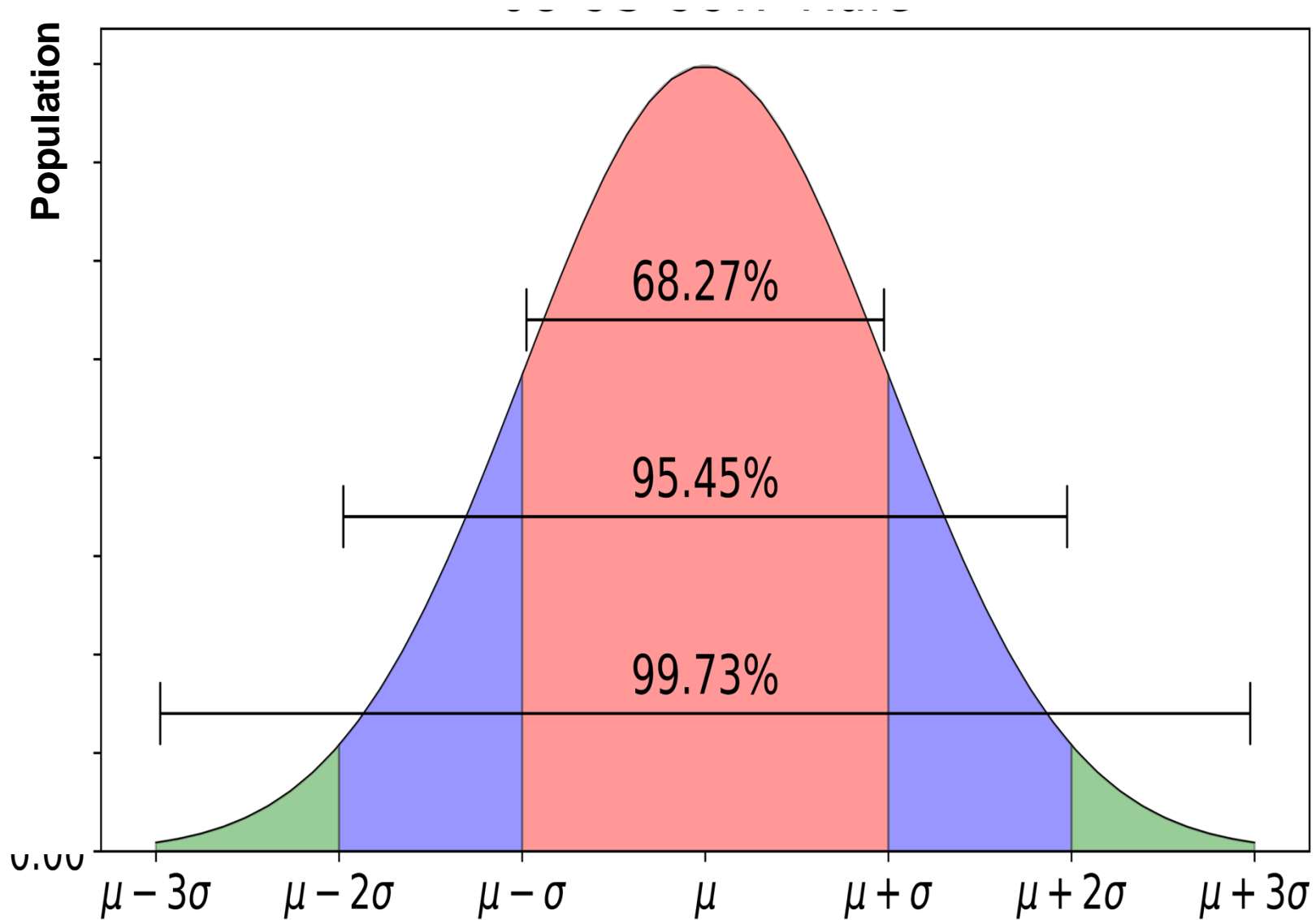


# La distribution normale de Gauss



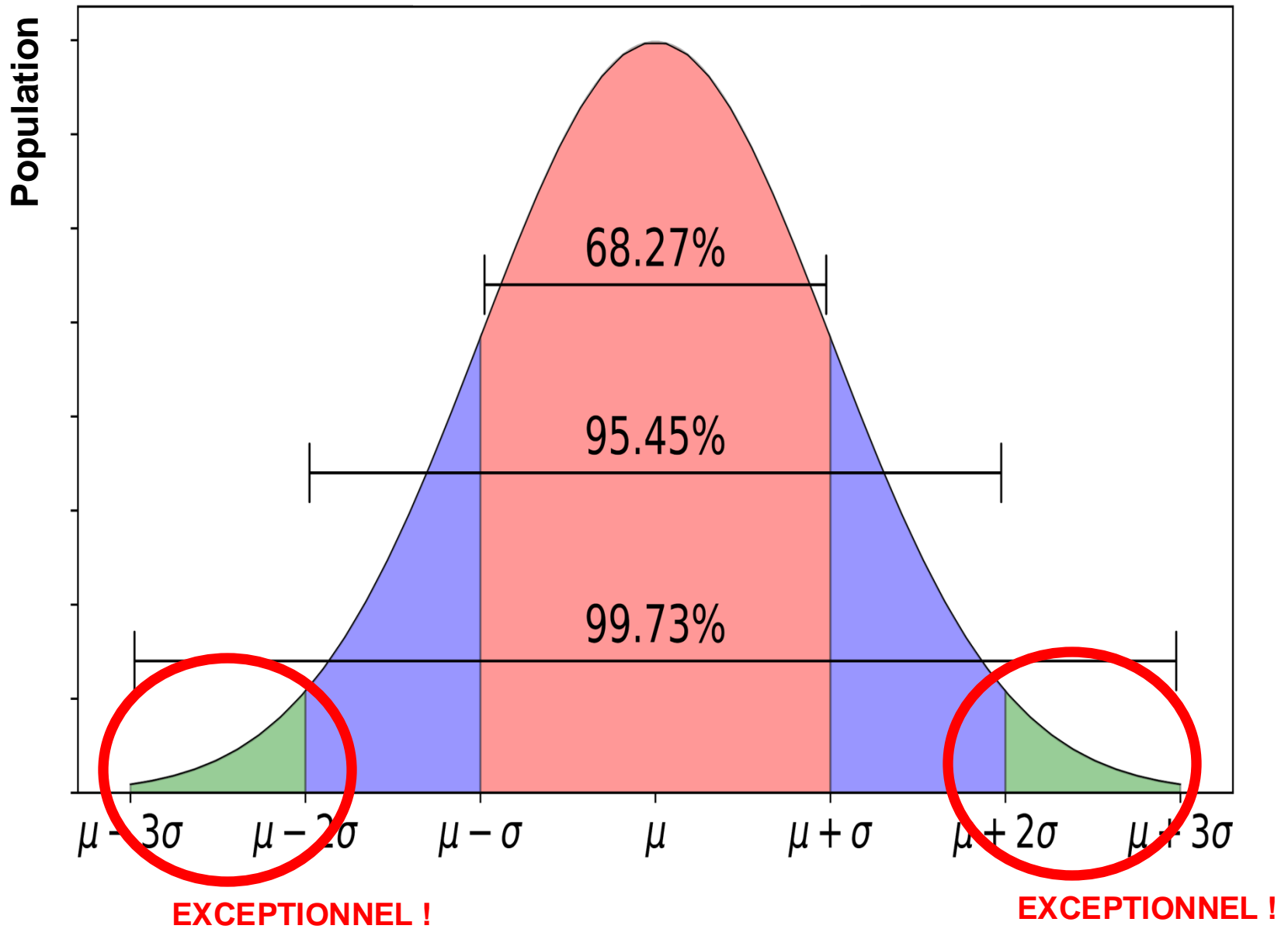


# La distribution normale de Gauss



95% de la population entre [moyenne - 2 écarts-types et moyenne + 2 écarts-types]  
=> C'est la population « normale ».

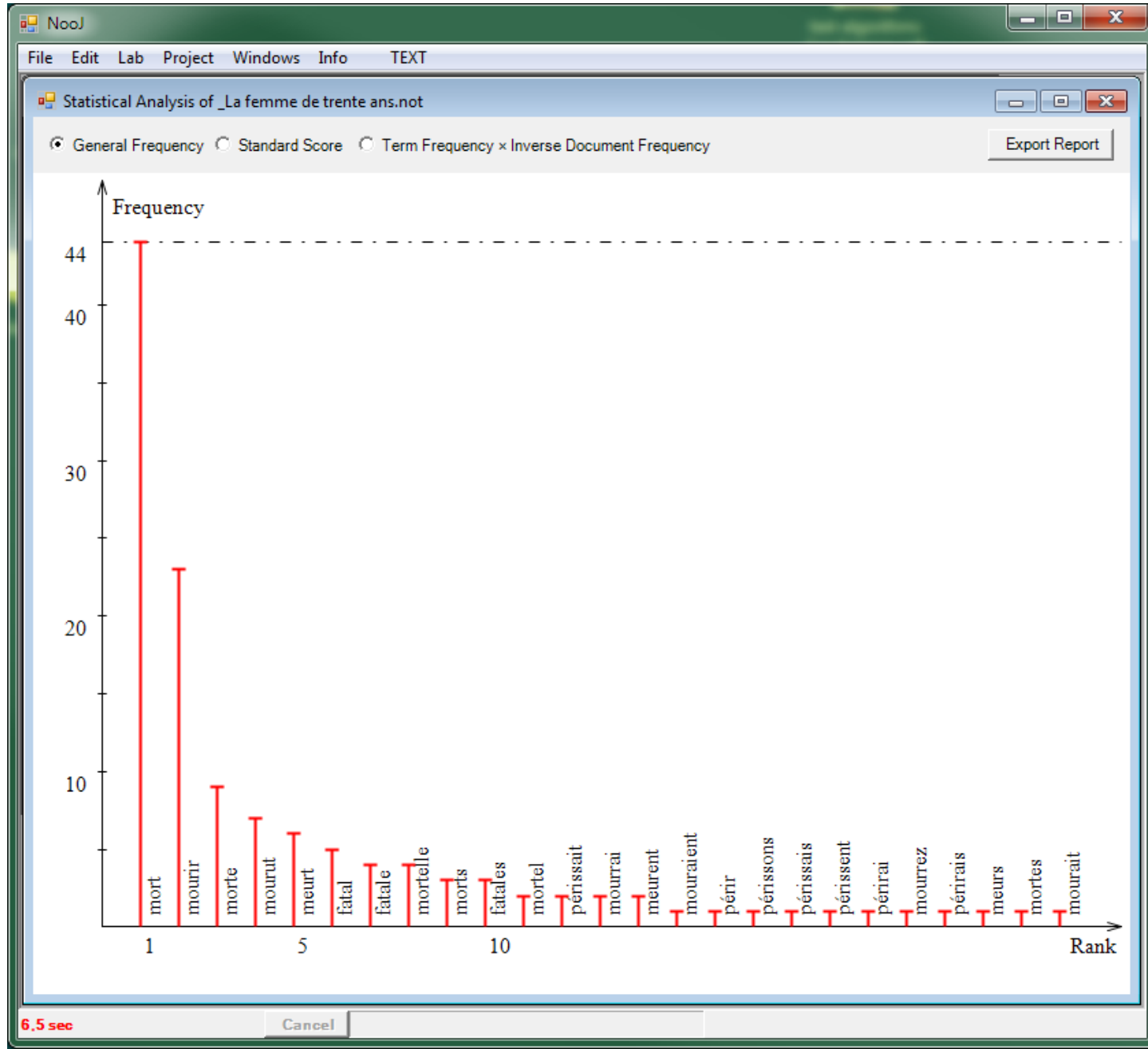
# La distribution normale de Gauss



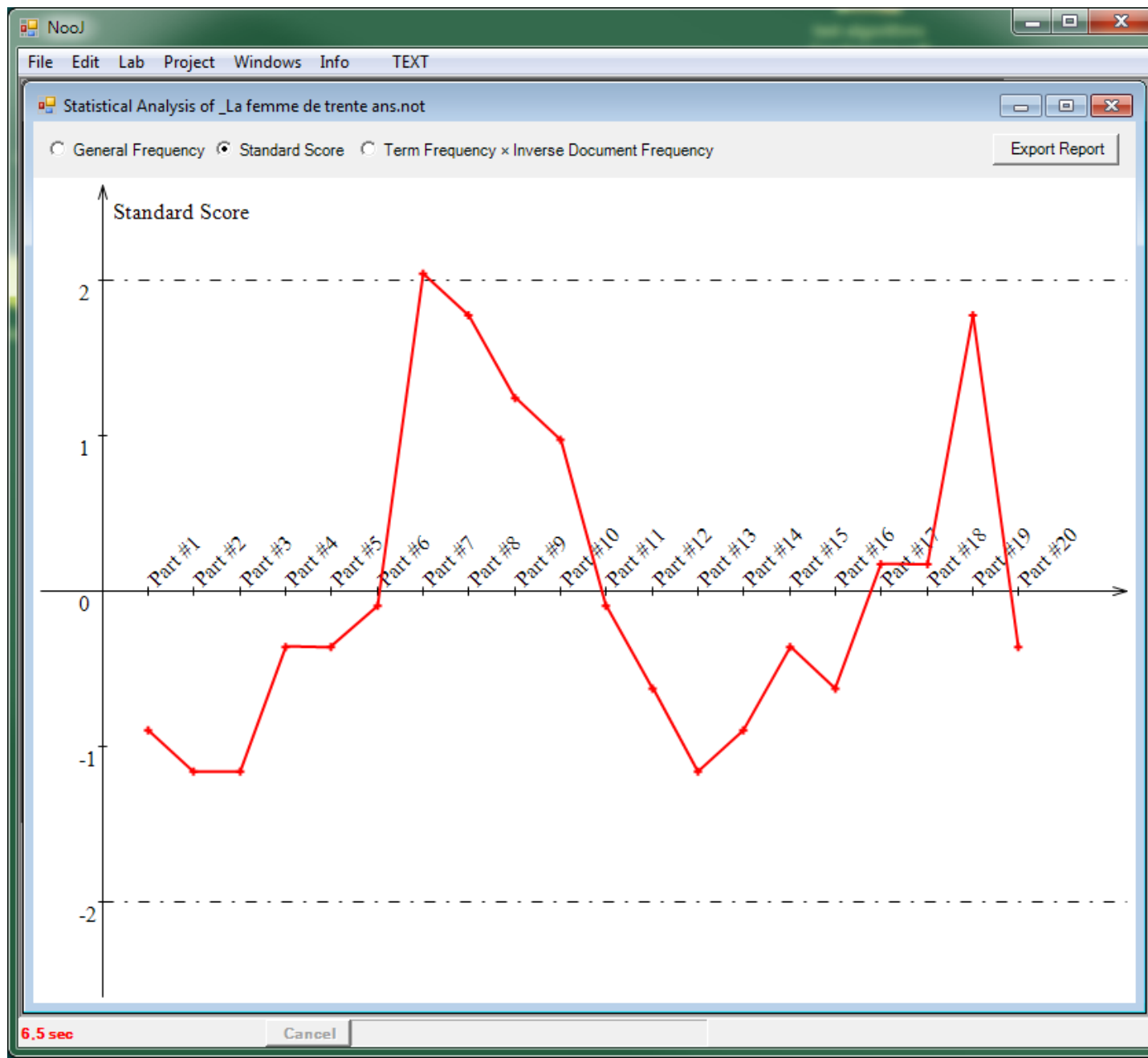
# Le score standard

- CONCORDANCE > Statistical Analysis > Standard Score
- Le score standard mesure la distance d'une mesure à la moyenne
- Si le score standard est  $> 2$ , alors la mesure est anormalement élevée  
=> très intéressant
- Si le score standard est  $< 2$ , alors la mesure est anormalement faible  
=> très intéressant
- Entre les deux, la mesure est "normale", i.e. inintéressant

# Analyse statistique d'une concordance : fréquence



# Analyse statistique d'une concordance : score standard



# Applications

- **Veille technologique.** On peut détecter la nouvelle utilisation d'un produit
- **Veille épidémiologique**, ex. CHU de Marseille. On peut détecter l'apparition de symptômes exceptionnels
- **Analyse psychologique**, ex. entretiens d'enfants. On peut détecter l'absence d'un thème
- **Veille sécuritaire, Veille marketing**, etc.

# Analyser l'importance d'une structure grammaticale

- Est-ce que le futur est utilisé par Balzac de façon anodine, ou alors d'une façon intéressante comme outil narratif ?
- **Exercice** : caractériser les phrases qui expriment le futur ; est-ce inintéressant (i.e., purement grammatical), ou est-ce pertinent d'un point de vue narratif ?

# Analyser l'importance d'une structure grammaticale

- Construire la requête pour identifier toutes les phrases qui expriment le futur



# Le futur

- Futur conjugué : toutes les formes verbales qui ont la propriété “+F” (futur de l’indicatif)

<V+F>

# Grammaire du futur

- Ajouter le futur proche : *aller* suivi d'un verbe à l'infinitif

<V+F> |

<aller> <V+INF>

# Grammaire du futur

- Attention : *aller* doit être conjugué au présent (pour éviter *nous allons déjeuner*)

<V+F> |

<aller+PR> <V+INF>

# Grammaire du futur

- On veut aussi les phrases négatives :

<V+F> |

<aller+PR> (<E> | pas | jamais | guère) <V+INF>

# Grammaire du futur

- On peut trouver n'importe quel adverbe entre aller et le verbe à l'infinitif, pas simplement les négations, ex.:

*Il va bêtement recommencer*

<V+F> |

<aller+PR> (<E> | <ADV>) <V+INF>

# Grammaire du futur

- On peut trouver plusieurs adverbes, ex. :  
*Il ne va plus jamais fumer.*

<V+F> |

<aller+PR> <ADV>\* <V+INF>

# Grammaire du futur

- L'adverbe *demain* caractérise le futur, même dans une phrase au présent, ex. :

*Il n'y a pas cours demain.*

- Il y a d'autres expressions qui caractérisent une date future :

<V+F> |

<aller+PR> <ADV>\* <V+INF> |

demain |

(lundi|mardi|mercredi|jeudi|vendredi|samedi|dimanche) prochain |

(printemps | été | automne | hiver) prochain |

dans <NB> (secondes | minutes | heures | jours | semaines | mois | trimestres | ans | décennies | siècles | millénaires)

- **Exercice** : appliquer la requête au texte, afficher l'analyse par score standard. Est-ce que le futur est utilisé comme outil narratif ?