

Recherche et extraction d'information

CNRS TDM2021

max.silberztein@univ-fcomte.fr

1. Formes graphiques *vs.* Unités linguistiques

Recherche d'information : concepts ou formes graphiques ?

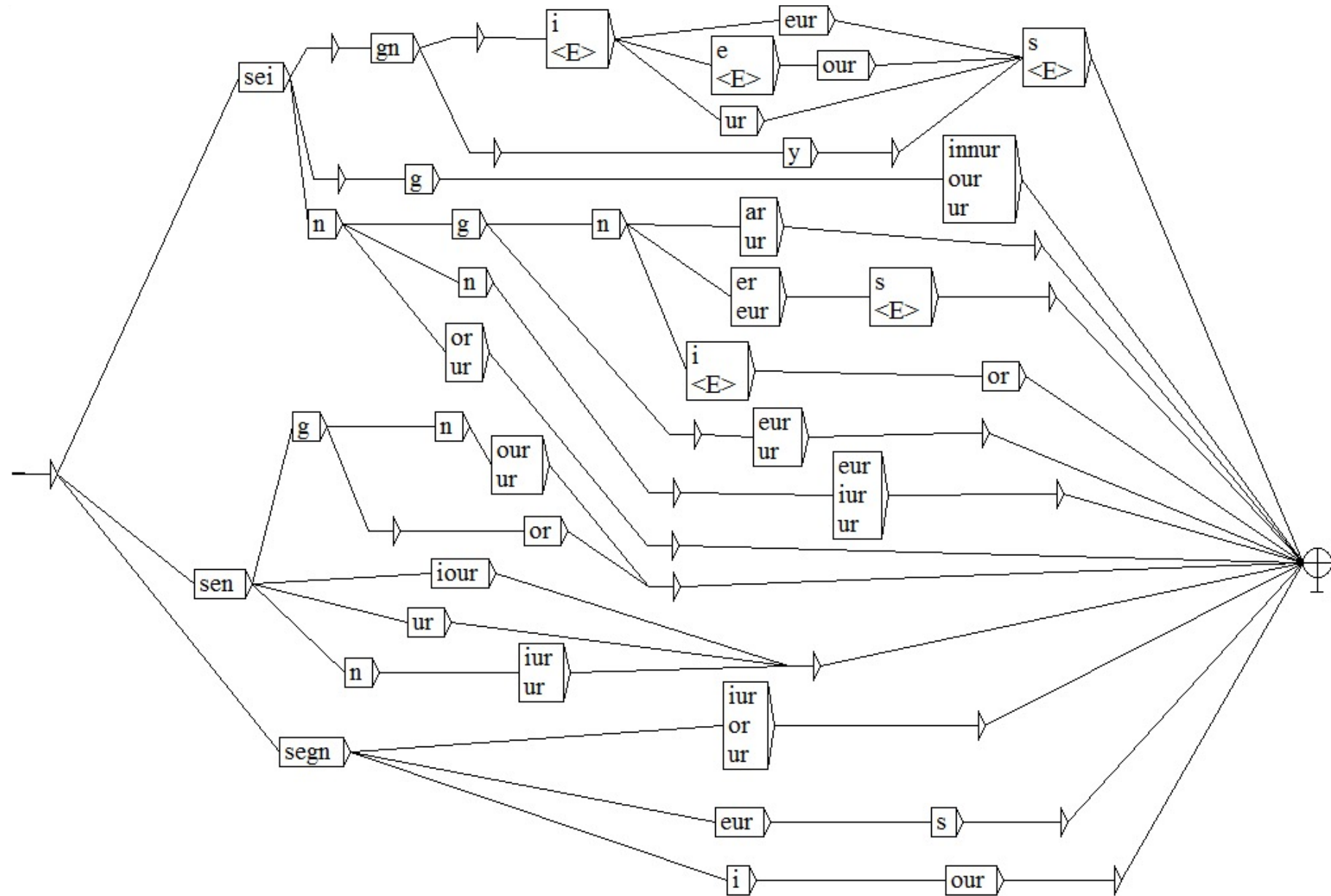
- On a besoin d'outils pour rechercher des informations dans les corpus
 - Les moteurs de recherche et outils d'exploration de texte utilisés : *Alceste, Google, Lexico, Iramuteq, Sketch Engine, Trameur, TXM*, etc.
- ...traitent **formes orthographiques** présentes dans les fichiers-textes

Formes orthographiques ↔ information

Formes graphiques → Norme orthographique

- L'orthographe n'a pas toujours été fixée, ex. corpus du XIVe siècle
- L'orthographe n'est pas vraiment présente dans les SMS, tweets, etc.
- Même dans les textes standard (articles, romans), l'orthographe n'est pas fiable, ex. *audiovisuel* (26 M) *audio visuel*, *audio-visuel* (1 M) ;
- Plus de 5% d'incohérences entre les dictionnaires de référence : (GDEL, GR, TLF) : *simulateur de vol* ou *simulateur de vols*, *Afrique Noire* ou *Afrique noire*, *czar* ou *tsar*, *Moyen Âge* ou *moyen Age*, etc.

Solution linguistique : décrire toutes les variantes



2. Lemmatisation

2. Lemmatisation ou non ?

- Certains outils (ex. *Lexico*) revendiquent la distinction entre les formes fléchies, ex. *liberté* vs. *libertés* n'ont pas le même sens
- En conséquence, ils traitent indépendamment les formes des termes, même lorsqu'il n'y a pas de différence de sens (autre que le nombre), ex. :
 - Une révolution → deux révolutions*
 - Une table → deux tables*
 - Une girafe → deux girafes*
 - Un boulanger → deux boulangers*
- Les fréquences de plus de 95% des termes sont alors sous-estimées → On perd de l'information

Lemmatisation par approche stochastique, ex. *Annie*, *Gate*, *NLTK*, *TreeTagger*, etc. utilisent des corpus d'apprentissage

- Les outils stochastiques de recherche d'information utilisent des étiqueteurs (taggers) stochastiques qui eux-mêmes utilisent des corpus d'apprentissage comme modèles pour analyser le texte à analyser
- Penn Treebank (Univ. de Pennsylvania) : 1 million de mots
- Open American National Corpus (OANC) : 11 millions de mots
- Corpus Of Contemporary American English (COCA) : 1,1 milliard de mots
- Corpus français (Univ. de Leipzig, Univ. de Neuchâtel) : 700 millions de mots
- Corpus Frantext (ATILF) : 262 millions de mots
- etc.

The Penn Treebank

Battle-tested/SingularProperName Japanese/SingularProperName
industrial/Adjective managers/PluralNoun here/Adverb always/Adverb
buck/Verb up/Preposition nervous/Adjective newcomers/PluralNoun
with/Preposition the/Determiner tale/SingularNoun of/Preposition
the/Determiner first/Adjective of/Preposition their/PossessivePronoun
countrymen/NomPluriel to/to visit/Verbe Mexico/NomPropreSingulier,/,
a/Determiner boatload/SingularNoun of/Preposition samurai/PluralNoun
warriors/PluralNoun blown/PastParticiple ashore/Adverb 375/Number
years/PluralNoun ago/Adverb ./.

From/Preposition the/Determiner beginning/SingularNoun ,/,
it/PersonalPronoun took/PastParticiple a/Determiner man/SingularNoun
with/Preposition extraordinary/Adjective qualities/PluralNoun to/TO
succeed/Verb in/Preposition Mexico/SingularProperName,/, " /"
says/VerbPresent3rdSingular Kimihide/SingularProperName Takimura/
SingularProperName ,/, president/SingularNoun of/Preposition
Mitsui/PluralNoun group/SingularNoun 's/Possessive
Kensetsu/SingularProperName Engineering/SingularProperName
Inc./SingularProperName unit/SingularNoun ./.

The Penn Treebank

Battle-tested/SingularProperName Japanese/SingularProperName
industrial/Adjective managers/PluralNoun here/Adverb always/Adverb
buck/Verb up/Preposition nervous/Adjective newcomers/PluralNoun
with/Preposition the/Determiner tale/SingularNoun of/Preposition
the/Determiner first/Adjective of/Preposition their/PossessivePronoun
countrymen/NomPluriel to/to visit/Verbe Mexico/NomPropreSingulier,/,
a/Determiner boatload/SingularNoun of/Preposition samurai/PluralNoun
warriors/PluralNoun blown/PastParticiple ashore/Adverb 375/Number
years/PluralNoun ago/Adverb ./.
From/Preposition the/Determiner beginning/SingularNoun ,/,
it/PersonalPronoun took/PastParticiple a/Determiner man/SingularNoun
with/Preposition extraordinary/Adjective qualities/PluralNoun to/TO
succeed/Verb in/Preposition Mexico/SingularProperName,/, " /"
says/VerbPresent3rdSingular Kimihide/SingularProperName Takimura/
SingularProperName ,/, president/SingularNoun of/Preposition
Mitsui/PluralNoun group/SingularNoun 's/Possessive
Kensetsu/SingularProperName Engineering/SingularProperName
Inc./SingularProperName unit/SingularNoun ./.

The Penn Treebank

Battle-tested/SingularProperName Japanese/SingularProperName

industrial/Adjective managers/PluralNoun here/Adverb always/Adverb
buck/Verb up/Preposition nervous/Adjective newcomers/PluralNoun
with/Preposition the/Determiner tale/SingularNoun of/Preposition
the/Determiner first/Adjective of/Preposition their/PossessivePronoun
countrymen/NomPluriel to/to visit/Verbe Mexico/NomPropreSingulier,/,
a/Determiner boatload/SingularNoun of/Preposition samurai/PluralNoun
warriors/PluralNoun blown/PastParticiple ashore/Adverb 375/Number
years/PluralNoun ago/Adverb ./.

From/Preposition the/Determiner beginning/SingularNoun ,/,
it/PersonalPronoun took/PastParticiple a/Determiner man/SingularNoun
with/Preposition extraordinary/Adjective qualities/PluralNoun to/TO
succeed/Verb in/Preposition Mexico/SingularProperName,/, " /"
says/VerbPresent3rdSingular Kimihide/SingularProperName Takimura/
SingularProperName ,/, president/SingularNoun of/Preposition
Mitsui/PluralNoun group/SingularNoun 's/Possessive
Kensetsu/SingularProperName Engineering/SingularProperName
Inc./SingularProperName unit/SingularNoun ./.

The Penn Treebank

Battle-tested/SingularProperName Japanese/SingularProperName industrial/Adjective managers/PluralNoun here/Adverb always/Adverb **buck/Verb up/Preposition** nervous/Adjective newcomers/PluralNoun with/Preposition the/Determiner tale/SingularNoun of/Preposition **the/Determiner first/Adjective of/Preposition their/PossessivePronoun** countrymen/NomPluriel to/to visit/Verbe Mexico/NomPropreSingulier,/, **a/Determiner boatload/SingularNoun of/Preposition samurai/PluralNoun warriors/PluralNoun blown/PastParticiple ashore/Adverb 375/Number years/PluralNoun ago/Adverb ./.**
From/Preposition the/Determiner beginning/SingularNoun-/, **it/PersonalPronoun took/PastParticiple** a/Determiner man/SingularNoun with/Preposition extraordinary/Adjective qualities/PluralNoun **to/TO** succeed/Verb in/Preposition Mexico/SingularProperName,/, " /" says/VerbPresent3rdSingular **Kimihide/SingularProperName Takimura/SingularProperName** ,/, president/SingularNoun of/Preposition **Mitsui/PluralNoun group/SingularNoun 's/Possessive Kensetsu/SingularProperName Engineering/SingularProperName Inc./SingularProperName** unit/SingularNoun ./.

Les corpus modernes : COCA, OANC, etc.

- on trouve 20% d'étiquettes impossibles, ex. :
 - abbreviate, abduct, abhor, abhors, etc.* (~~Noms~~),
 - about, agonized, bible, cactus, California, etc.* (~~Adjectifs~~)
 - expenditures, Japanese, many, initiatives, wimp, etc.* (~~Verbes~~)
 - anomaly, back, because, by, of, out, particular, upon, etc.*
(~~Adverbes~~)...
- Agglutations systématiquement étiquetées comme ~~Noms communs~~, ex. :
 - audienceless, autodialed, bioremediation, barklike, etc.*

Les fautes de frappe sont systématiquement étiquetées comme noms

- En minuscules -> ~~Noms communs~~
*absentionists (NNS), achives (NNS), accrossthe (NN),
afteryou (NN), alwaysattend (NN) ...*
- En majuscule -> ~~Noms propres~~:
Aconfession, Afew, AffairsThe, Allpolitics...

Les mots composés ne sont pas traités correctement

- Séquences étiquetées incorrectement comme des mots composés :

*abide-and, abuse-suggesting, activity-regarded,
adoption-related, Afghan-based*

- Mots composés ignorés :

As_IN a_DT matter_NN of_IN fact_NN

- Incohérences :

*left-wing, wing-feathers and ultra-left-wing
left wing, wing commander, wing nuts*

CONCORDANCE

French Web 2017 (frT...)

Obtenez plus d'espace +

lemme **aimer** 1871115

filtre **Arranet 1, 2** 23 inférieur à 0.01 X

phrase

Détails

phrase

1 [doc#15462](#)

<S> C' est d' ailleurs au cours d' un
PP3MSN0/ce VMIP3S0/être SP/de RG/ailleurs SP/à NCFP000/cour SP/de DI0MS0/un
voyage à Versailles qu' il rencontre **Aimée**
NCMS000/voyage SP/à NPFP000/versailles PR00000/que PP3MSN0/il VMIP1S0/rencontrer VMP00SF/aimer
d' **Arranet** , dont il tombe éperdument
SP/de NPMS000/arranet Fc/, PR00000/dont PP3MSN0/il VMIP1S0/tomber RG/éperdument
amoureux dès le premier regard . </S>
AQ0MN00/amoureux SP/dès DA0MS0/le AQ0MS00/premier NCMS000/regard Fp/.

2 [doc#15462](#)

<S> Il doute aussi de son amour pour **Aimée**
PP3MSN0/il VMIP1S0/douter RG/aussi SP/de DP3CSS/son NCMS000/amour SP/pour VMP00SF/aimer
d' **Arranet** et pense de plus en plus à la
SP/de NPMS000/arranet CC/et VMIP1S0/penser SP/de RG/plus SP/en RG/plus SP/à DA0FS0/le
Satin : " A l' amour , feu couvant
NPMS000/satin Fd/: Fe/ VMIP3S0/avoir DA0CS0/le NCMS000/amour Fc/, NCMS000/feu VMG0000/couver
, qu' il continuait à lui porter , s'
Fc/, CS/que PP3MSN0/il VMIP1S0/continuer SP/à PP3CS00/lui VMN0000/porter Fc/, PP3CN00/se
ajoutaient une estime , un respect et une



CONCORDANCE

French Web 2017 (frT...

Obtenez plus d'espace

lemme **aimer** 1871115

filtre **Arranet 1, 2** 23 inférieur à 0.01

phrase

Détails

phrase

1 [doc#15462](#)

<S> C' est d' ailleurs au cours d' un
 PP3MSN0/il ~~VMIP1S0/doute~~ RG/ailleurs SP/à ~~NCFP000/ce~~ SP/de DI0MS0/un
 voyage à Versailles qu' il rencontre **Aimée**
 NCMS000/voyage SP/à NPFP000/versailles PR00000/que PP3MSN0/il ~~VMIP1S0/rencontre~~ ~~VMIP000CF/aimer~~
 d' **Arranet** , dont il tombe éperdument
 SP/de NPMS000/arranet Fc/, PR00000/dont PP3MSN0/il ~~VMIP1S0/tomber~~ RG/éperdument
 amoureux dès le premier regard . </S>
 AQ0MN00/amoureux SP/dès DA0MS0/le AQ0MS00/premier NCMS000/regard Fp/.

2 [doc#15462](#)

<S> Il doute aussi de son amour pour **Aimée**
 PP3MSN0/il ~~VMIP1S0/doute~~ RG/aussi SP/de DP3CSS/son NCMS000/amour SP/pour ~~VMIP000CF/aimer~~
 d' **Arranet** et pense de plus en plus à la
 SP/de NPMS000/arranet CC/et ~~VMIP1S0/pense~~ SP/de RG/plus SP/en RG/plus SP/à DA0FS0/le
 Satin : " A l' amour , feu couvant
 NPMS000/satin Fd/: Fe/ ~~VMIP000/ave~~ DA0CS0/le NCMS000/amour Fc/, NCMS000/feu VMG0000/couver
 , qu' il continuait à lui porter , s'
 Fc/, CS/que PP3MSN0/il VMII3S0/continuer SP/à PP3CS00/lui VMN0000/porter Fc/, PP3CN00/se
 ajoutaient une estime , un respect et une



CONCORDANCE

French Web 2017 (frT...)

Obtenez plus d'espace +

lemme **aimer** 1871115

filtre **Arranet 1, 2** 23 inférieur à 0.01

phrase

Détails

phrase

1 [doc#15462](#)

<S> C' est d' ailleurs au cours d' un
PPSCNN0/ce VMIP3S0/être SP/de RG/ailleurs SP/à NCFS000/cour SP/de DI0MS0/un
voyage à Versailles qu' il rencontre **Aimée**
NCMS000/voyage SP/à NPFP000/versailles PR00000/que PP3MSN0/il VMIP1S0/rencontre VMPC00SF/aimer
d' **Arranet** , dont il tombe éperdument
SP/de NPMS000/arranet Fc/, PR00000/dont PP3MSN0/il VMIP1S0/tomber RG/éperdument
amoureux dès le premier regard . </S>
AQ0MN00/amoureux SP/dès DA0MS0/le AQ0MS00/premier NCMS000/regard Fp/.

2 [doc#15462](#)

<S> Il doute aussi de son amour pour **Aimée**
PP3MSN0/il VMIP1S0/douter RG/aussi SP/de DP3CSS/son NCMS000/amour SP/pour VMPC00SF/aimer
d' **Arranet** et pense de plus en plus à la
SP/de NPMS000/arranet CC/et VMIP1S0/penser SP/de RG/plus SP/en RG/plus SP/à DA0FS0/le
Satin : " A l' amour , feu couvant
NPMS000/satin Fd/: Fe/ VMIP000/ave/ DA0CS0/le NCMS000/amour Fc/, NCMS000/feu VMG0000/couver
, qu' il continuait à lui porter , s'
Fc/, CS/que PP3MSN0/il VMII3S0/continuer SP/à PP3CS00/lui VMN0000/porter Fc/, PP3CN00/se
ajoutaient une estime , un respect et une

Solution linguistique pour lemmatiser les termes

- Dictionnaire avec description flexionnelle, ex. :

liberté ,NOM+Abstrait+féminin+singulier+DROITE

libertés ,NOM+Abstrait+féminin+pluriel+GAUCHE

révolution ,NOM+Abstrait+féminin+FLX=TABLE

Solution linguistique pour lemmatiser les termes

- Dictionnaire avec description flexionnelle, ex. :

liberté, NOM+Abstrait+féminin+singulier

libertés, NOM+Abstrait+féminin+pluriel

révolution, NOM+Abstrait+féminin+**FLX=TABLE**

TABLE = <E>/singulier | s/pluriel

3. Dérivation

Exemple d'utilisation : on analyse la paix sociale dans un pays. Examinons dans les archives d'un quotidien la fréquence d'occurrence du terme *demonstration*

... **Demonstrators** shouted against police brutality and racism and some held signs against discrimination ...

<https://www.aa.com.tr/en/europe/demonstrators-in-france-protest-george-floyd-s-death/1867169>

... Wearing face masks, waving black flags and keeping two yards apart, thousands of Israelis **demonstrated** against prime minister Benjamin Netanyahu ...

<https://www.reuters.com/article/us-health-coronavirus-israel-protest/anti-netanyahu-rally-draws-thousands-under-coronavirus-curbs-idUSKBN2210S4>

... Parents to **redemonstrate** against air pollution, demanding officials to take immediate and efficient actions and for the well-being of children...

<https://mongolia.gogo.mn/r/156657>

Comment reconnaître les formes dérivées ?

Requête « amour »

amour, amouracher, amoureuse, amoureusement

mais aussi :

~~amphithéâtre, ample, amusa, amusant...~~

et non :

aimer, aimait, aimons etc.

The screenshot shows a software interface with a title bar 'C:\HYPERBAS\Rogon.tbk'. The main window is divided into three panes. The left pane contains instructions: 'Les sous-fréquences dans les textes sont réparties en binomes, dont le premier élément indique le numéro d'ordre du texte, le second la fréquence du mot'. Below this is a green button 'Cliquer sur un mot pour l'activer' and several navigation icons. The middle pane, titled 'Formes', lists the following words and their frequencies: 111 amour, 1 amour», 1 amouracher, 26 amoureuse, 8 amoureusement, 2 amoureuses, 55 amoureux, 49 amours, 1 amphigourique, 2 amphithéâtre, 1 ample, 5 ampleur, 1 amplifia, 13 amusa, 5 amusaient, 25 amusait, 16 amusant, 2 amusante, 1 amusantes. The right pane, titled 'Lemmes', shows 'amour' with a frequency of 111. Below it, there is a note: 'CLIQUER SUR UN TEXTE (ou sur TOUS LES TEXTES) pour afficher les contextes du mot "amour". Cliquer AILLEURS dans cette fenêtre.' At the bottom right, there is a list of other words: 3 and_3, 8 andouille_2, 1 andré_2.

Solution linguistique pour traiter la dérivation

- Dictionnaire avec description dérivationnelle, ex. :

manifeste , VERBE+tr+DRV=ANT:COUSIN+DRV=ATION:TABLE

ANT = <B2>/NOM

COUSIN = <E>/mas+sing | s/mas+plur | e/fem+sing | es/fem+plur

ATION = <B2>ation/NOM

TABLE = <E>/singulier | s/pluriel

3. Champs lexicaux

Champs lexicaux & nuages de mots

cook **verb** switch to [cook](#) (noun)

eat bake boil roast dry smoke wash fry consume pack clean mix taste chop drink feed
freeze slice prepare sell serve steam buy love try harvest store work like cut put soak
stuff heat wrap make enjoy cool live get



A word cloud visualization of the semantic field for the verb 'cook'. The most prominent words are 'eat', 'bake', 'boil', and 'roast'. Other significant words include 'wash', 'fry', 'consume', 'pack', 'clean', 'mix', 'taste', 'chop', 'drink', 'feed', 'freeze', 'slice', 'prepare', 'sell', 'serve', 'steam', 'buy', 'love', 'try', 'harvest', 'store', 'work', 'like', 'cut', 'put', 'soak', 'stuff', 'heat', 'wrap', 'make', 'enjoy', 'cool', 'live', and 'get'. The words are arranged in a circular pattern, with 'eat' and 'bake' being the largest and most central.

Requête « cook »

eat, bake, boil, roast

mais aussi :

~~*buy, enjoy, like, live, make, put, store, try...*~~

et non :

dish, meal, restaurant, oven, etc.

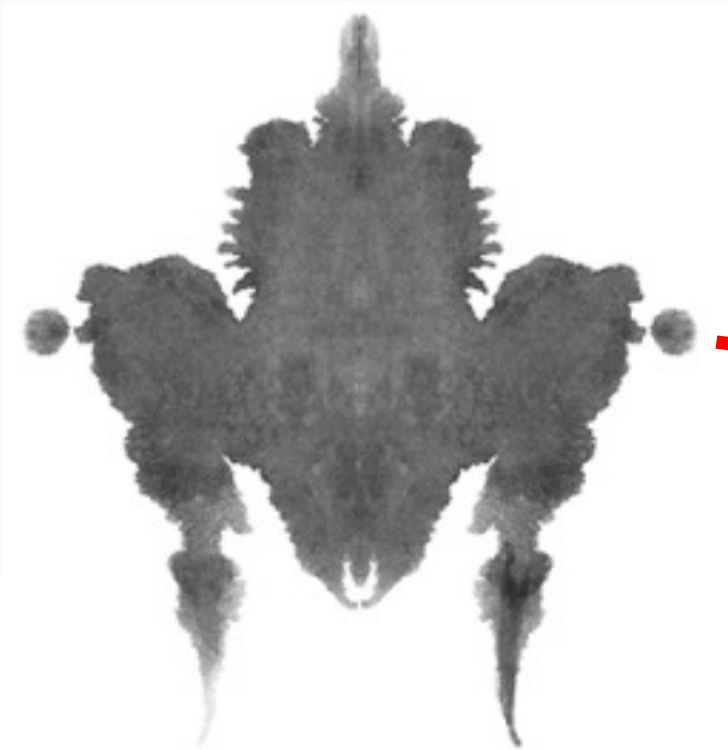
Champs lexicaux & nuages de mots

cook **verb** switch to cook (noun)

eat bake boil roast dry smoke wash fry consume pack clean mix taste chop drink feed

freeze slice prepare sell serve steam buy love try harvest store work like cut put soak

stuff heat wrap make enjoy cool live get



Le cerveau humain est doué pour inventer des connections



Requête « cook »

eat, bake, boil, roast

mais aussi :

buy, enjoy, like, live, make, put, store, try...

et non :

dish, meal, restaurant, oven, etc.

Champs lexicaux & nuages de mots

- Solution linguistique, adaptée à chaque utilisateur (policier ? psychologue ? littéraire ? médecin ?) :

MORT = <assassiner> | <clamcer> | <décéder>
| <égorger> | <massacrer> | <mourir> | <périr> |
<succomber> | <suicider> | <trépasser> | <tuer> |

<assassin> | <assassinat> | <cimetière> | funérailles
| <meurtrier> | <mort> | <veuf> |
<DET> (<A> | <E>) (tombe|tombes) |

<fatal> | <funeste> | <morte |
fatalement | funestement | mortellement ;

Champs lexicaux & nuages de mots

- Solution linguistique :

MORT = <assassiner> | <clamcer> | <décéder>
| <égorger> | <massacrer> | <mourir> | <périr> |
<succomber> | <suicider> | <trépasser> | <tuer> |

<assassin> | <assassinat> | <cimetière> | funérailles
| <meurtrier> | <mort> | <veuf> |

<DET> (<A> | <E>) (tombe | tombes) |

<fatal> | <funeste> | <morte |
fatalement | funestement | mortellement ;

Retrouver les occurrences du champ lexical MORT

- Créer une grammaire textuelle (rule-based) : File > New > Grammar
- Copier/coller l'expression après `Main =`
- Ne pas oublier de terminer l'expression par `;`
- Sauvegarder la grammaire
- Ouvrir le texte à traiter : File > Open > Text
- Appliquer la grammaire : TEXT > Locate > NooJ Grammar

Analyse d'opinion

- Rechercher des termes négatifs : les méthodes stochastiques (utilisant des corpus annotés et des probabilités calculées)
- **SentiWordNet** a été contruit à l'aide de méthodes stochastiques. Mais il comporte de nombreuses erreurs, ex. :
 - desillusion* et *uninviting* : polarité positive (PosScore = 0,75)
 - protect* et *soundproof* : polarité négative (NegScore = 0,75).

Analyse d'opinion : solution linguistique

- on construit un dictionnaire de termes négatifs ; on n'y insère aucune faute (ou alors on les corrige), ex. :

agressif,A+NEG+FLX=Actif

brutalement,ADV+NEG

détester,V+NEG+FLX=Aimer

- on construit des règles morphologiques qui transforment un terme neutre en terme négatif, ex. *bleu/bleuâtre*) :

Atre = hâtre/NEG ;

Ard = <B3>ard/NEG ;

Aille = aille/NEG ;

Asse = asse/NEG ;

- on construit des grammaires syntaxiques pour traiter la négation qui plonge un terme positif dans une expression négative, ex. *il ne gagnera plus jamais*.

Grammaire morphologique

NooJ - [Concordance]

File Edit Lab Project Windows Info Max TEXT CONCORDANCE

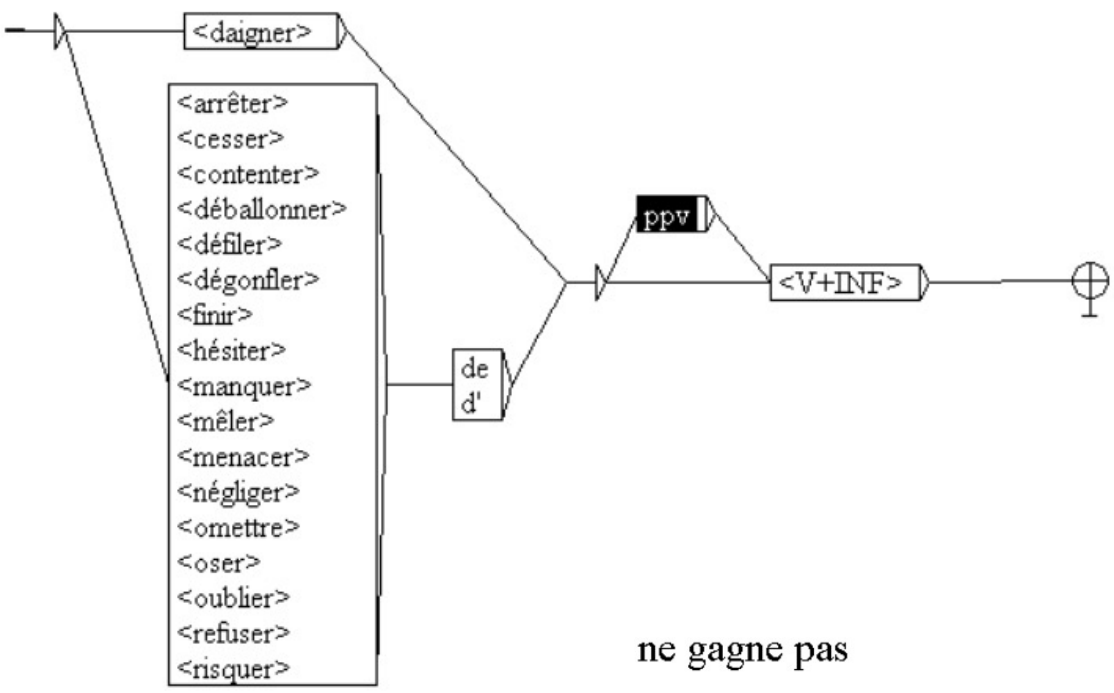
Reset Display: 5 characters before, and 5 after. Display: Matches Outputs word forms

Text	Before	Seq.	After
mauvaises pensées involontaires q...	imparfaite		nature engendre, mais qui montrent
un léger nuage qui dérobe	imparfaitement		le soleil. Vandenesse sortit après
petite main, bien gantée, froissait	impatiemment		un mouchoir, et ressemblait à
abîmes sans fond ! Elle attendait	impatiemment		le lever de sa fille
par le désir et l'	impatience		qui pétillaient dans tous les
que moqueur les signes d'	impatience		et de crainte qui se
tête agitée exprimait une extrême	impatience		, les deux pieds de devant
Julie en affectant de l'	impatience		. Puis, pour n'éveiller aucun
-elle avec une sorte d'	impatience		. Mais où sont pour moi
marquise l'attendit avec une	impatience		pleine de remords. Ecrire était
précieux, arrachaient des mouvem...	impatience		à la jolie femme: on
de Vandenesse qui partageait son	impatience		, ou le notaire qui plombait
'une fortune considérable. Dans son	impatience		de revoir son pays et
roula plus; bientôt un militaire	impatient		en ouvrit lui-même la
si beau que chaque notable,	impatient		de savoir s'il était
belle matinée, quelques négociants ...	impatients		de revenir dans leur patrie
l'accomplissement de la plus	imprudente		promesse que puissent se faire
remplir une mission dangereuse, m...	incapable		d'un commandement de quelque
mémoire à lui. Telle femme	incapable		de se rappeler les événements
Le notaire, l'imperturbable notaire,	incapable		de se demander pourquoi Mme

Query 857/857

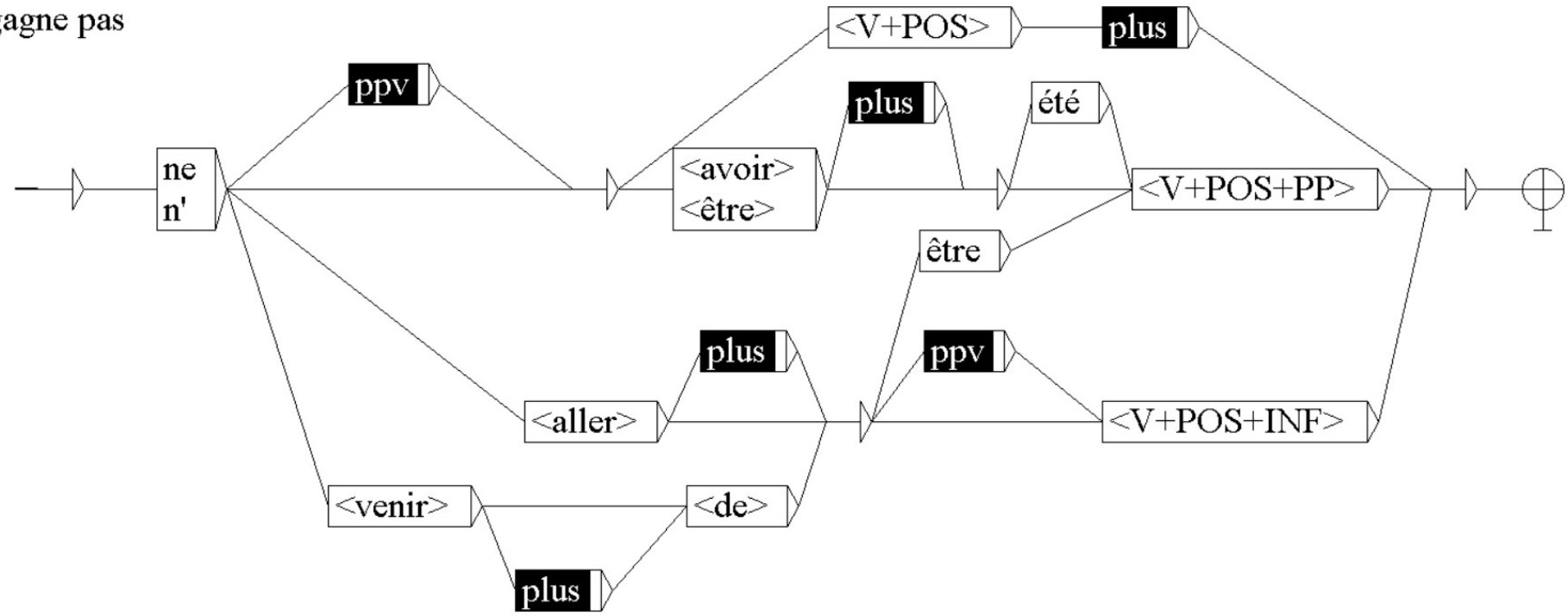
0,7 sec Cancel

Grammaire syntaxique



Verbes auxiliares, LG1

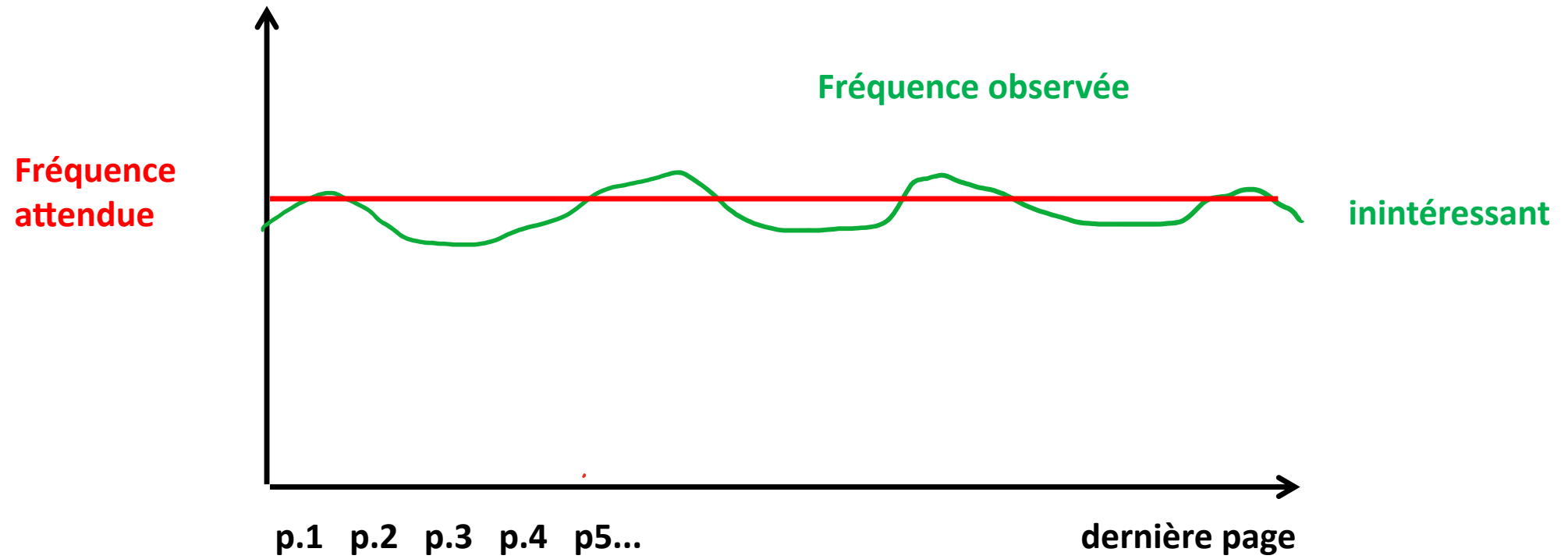
ne gagne pas



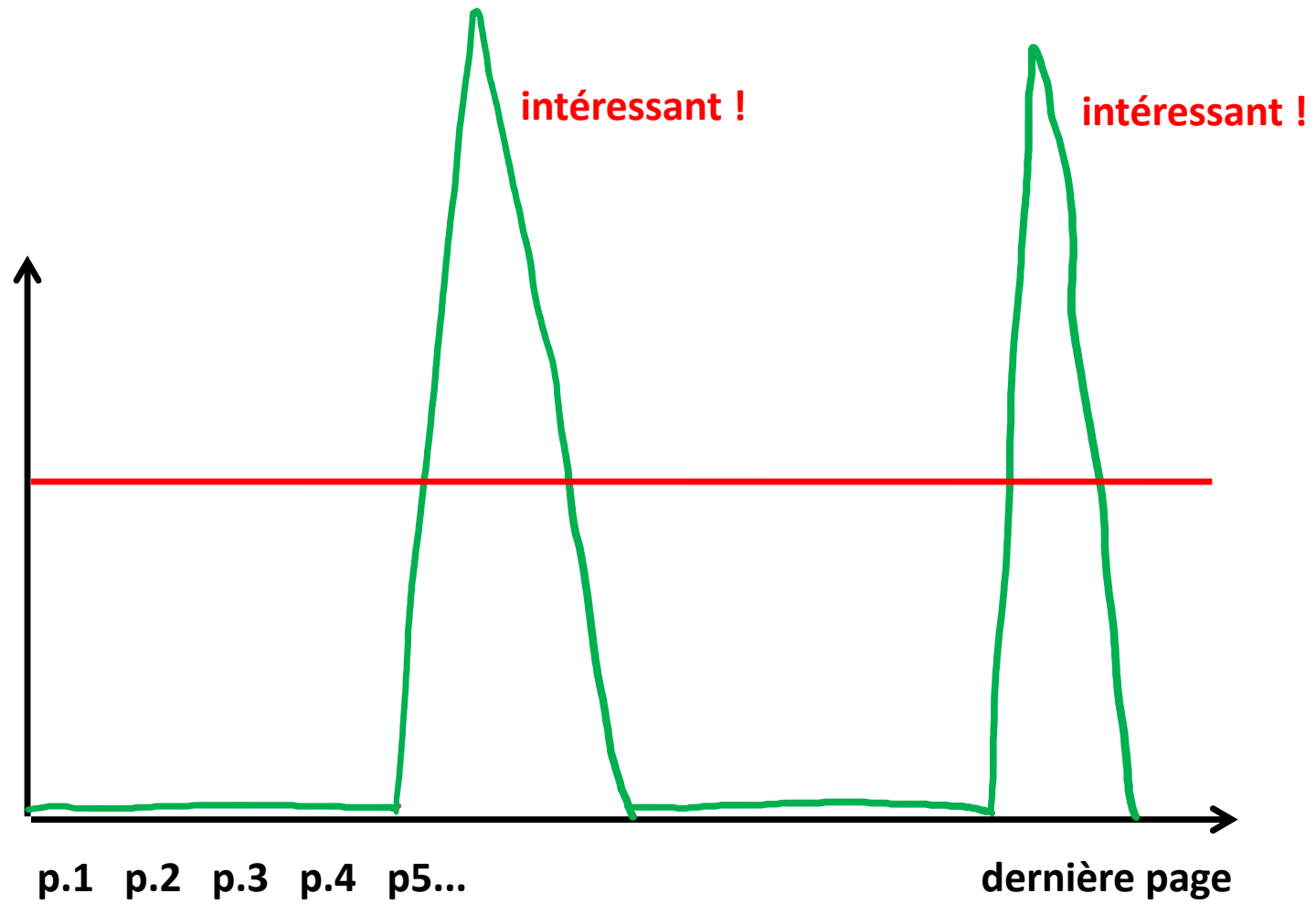
4. Analyses statistiques

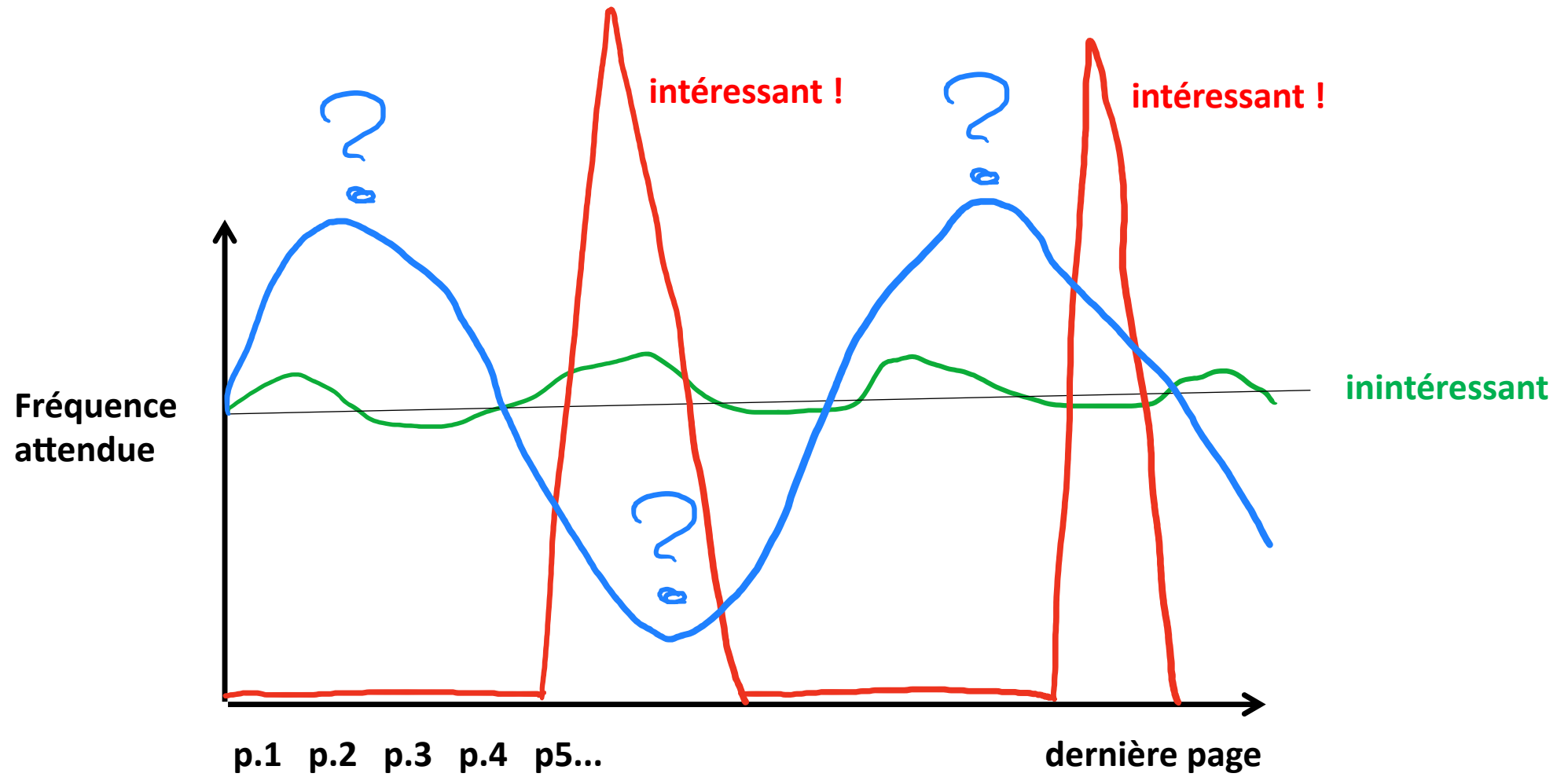
Analyse statistique

- 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 observée

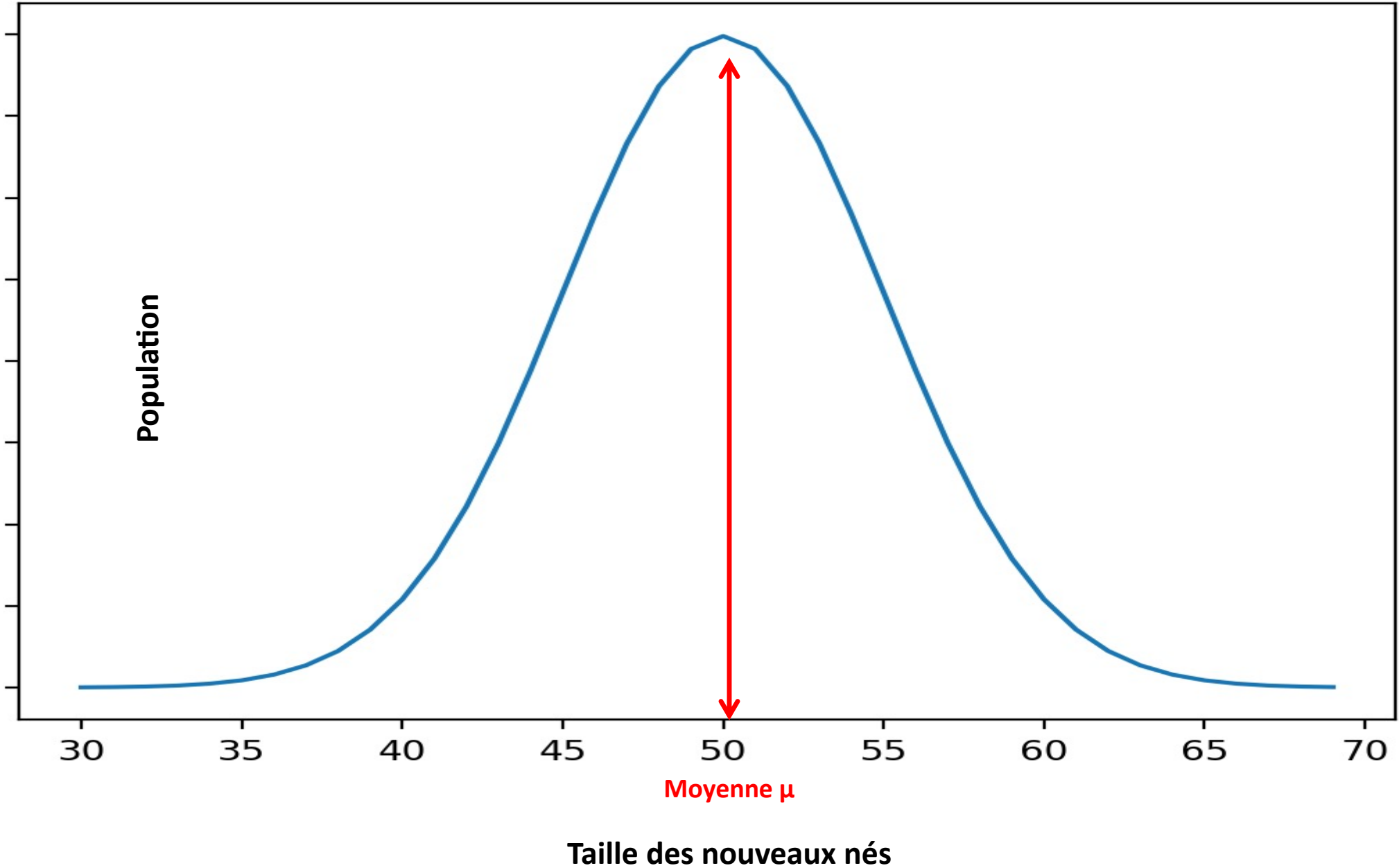


Fréquence
attendue

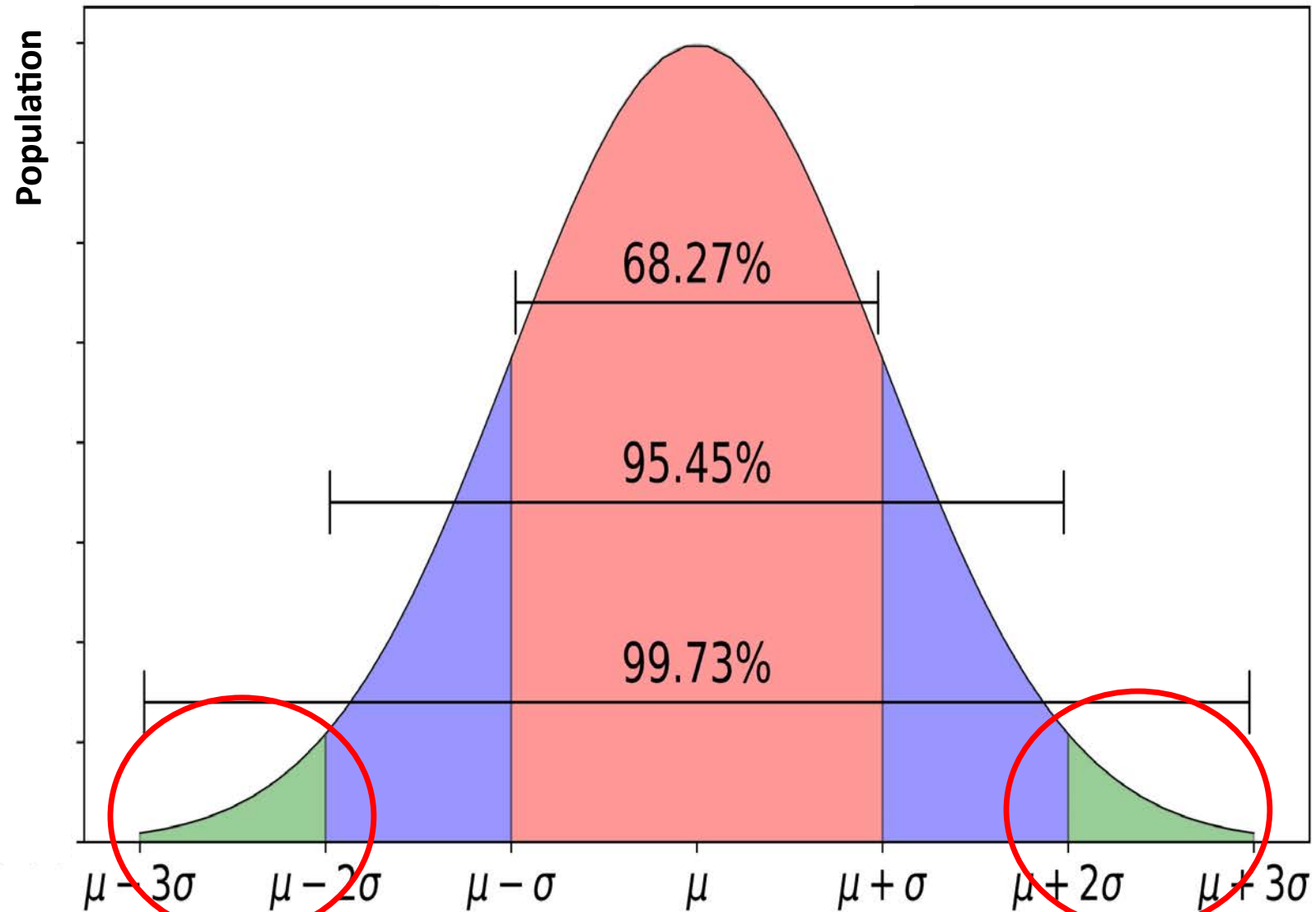




La distribution normale de Gauss



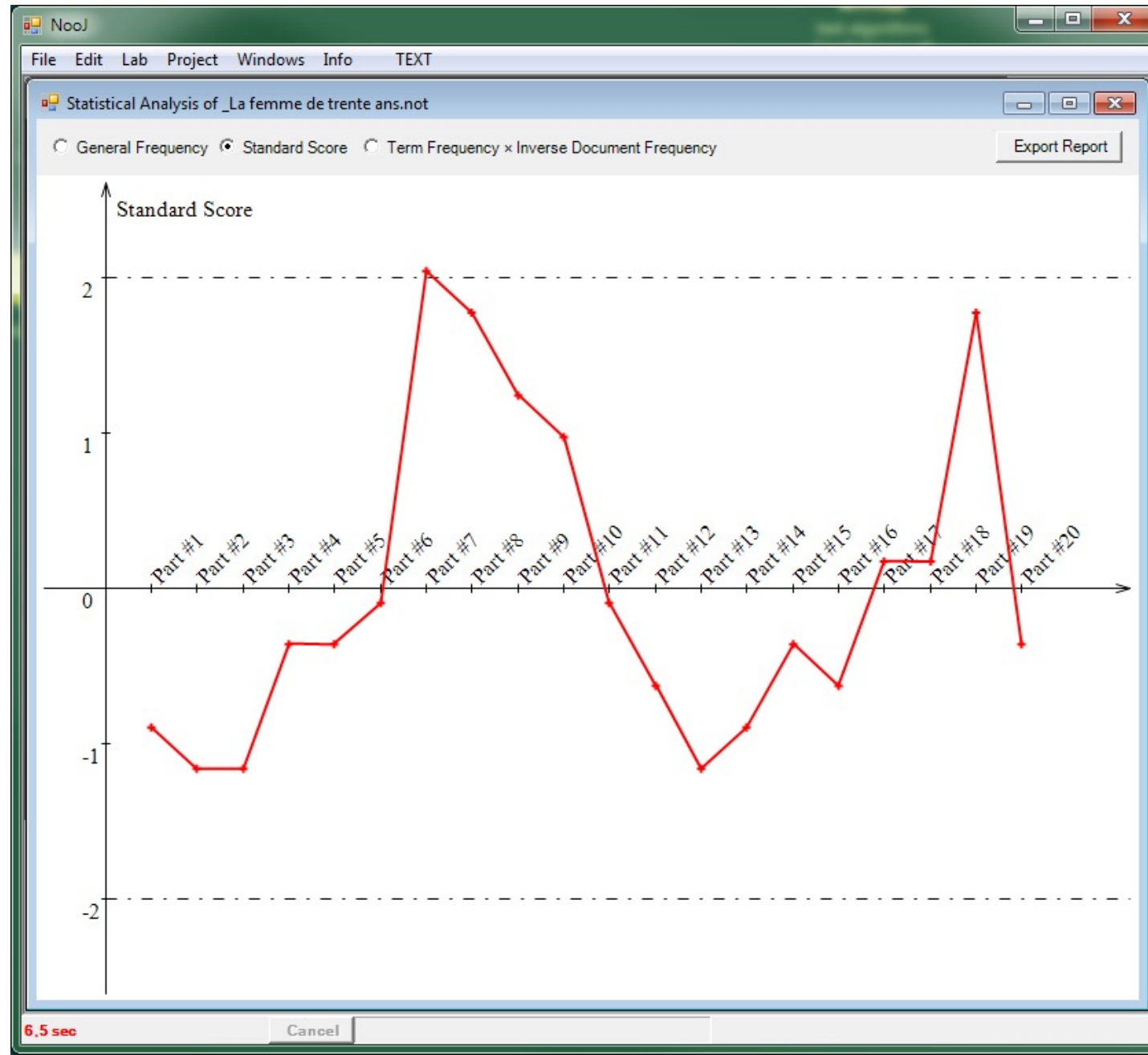
La distribution normale de Gauss



EXCEPTIONNEL !

EXCEPTIONNEL !

Analyse statistique d'une concordance : score standard



Applications

- Veille technologique, assurances, finance
- Veille épidémiologique, détection de l'arrivée du SRAS au Canada
- Mission MARS500
- Analyse de la négation dans les livres d'histoire donnés aux enfants hongrois, de 1870 à nos jours
- Analyse psychologique, ex. entretiens non directifs d'enfants
- etc.

Analyser l'importance de l'utilisation du futur

- Est-ce que le futur est utilisé dans le roman de façon grammaticale, anodine, ou alors d'une façon intéressante comme outil narratif ?
- On va chercher les phrases au futur ; on va analyser son score standard

Analyser l'importance de l'utilisation du futur

- Chercher toutes les phrases qui expriment le futur :

<V+F>

Analyser l'importance de l'utilisation du futur

- Chercher toutes les phrases qui expriment le futur :

<V+F> |

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

Analyser l'importance de l'utilisation du futur

- Chercher toutes les phrases qui expriment le futur :

<V+F> |

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

demain |

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

le (mois|trimestre|semestre|siècle) prochain |

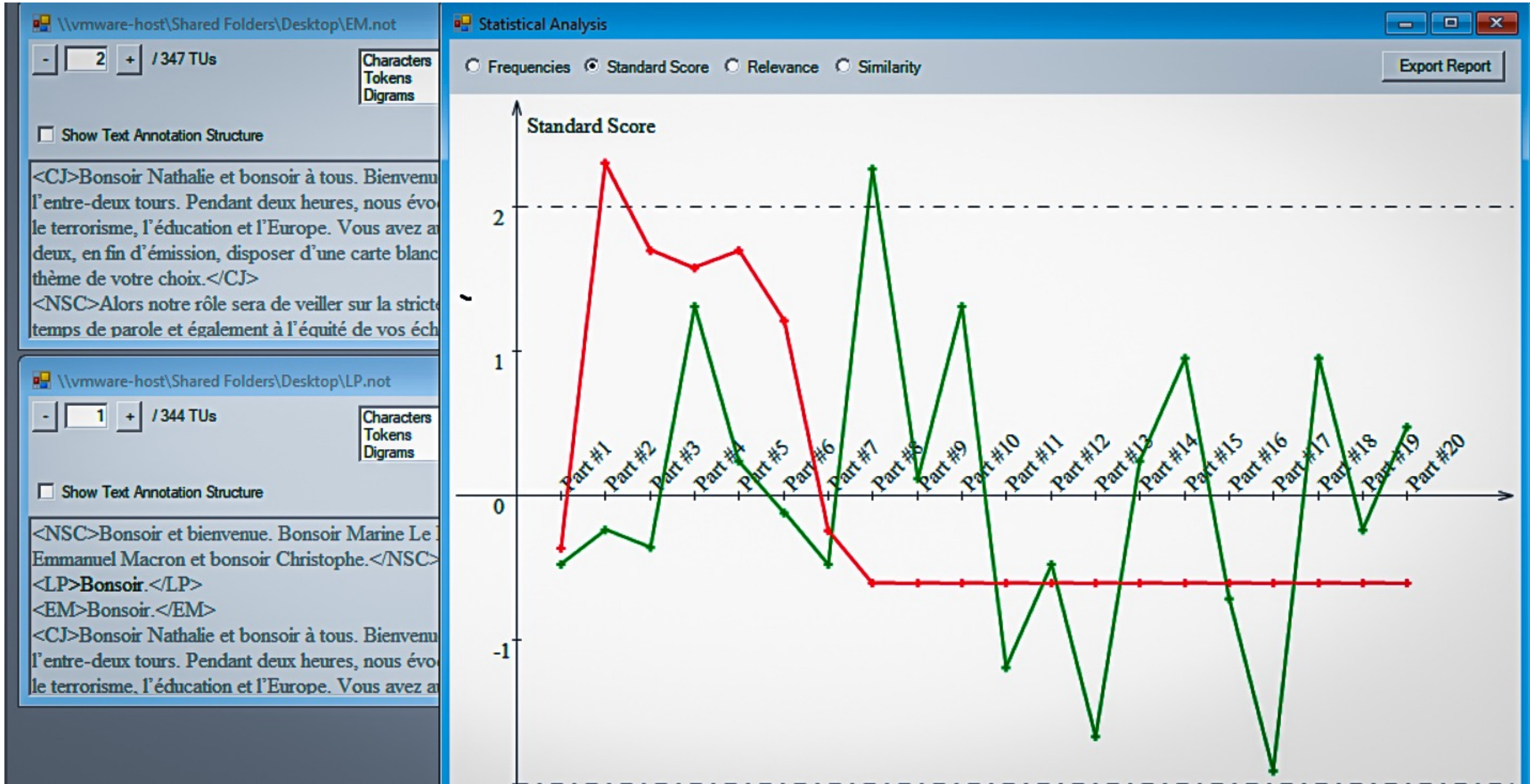
la (semaine|décennie) prochaine |

l'an prochain | l'année prochaine |

dans <NB> (jours|semaines|mois|ans|décennies|siècles)

Débat Macron-Le Pen, 2017

Termes et expressions négatives utilisé par chaque interlocuteur

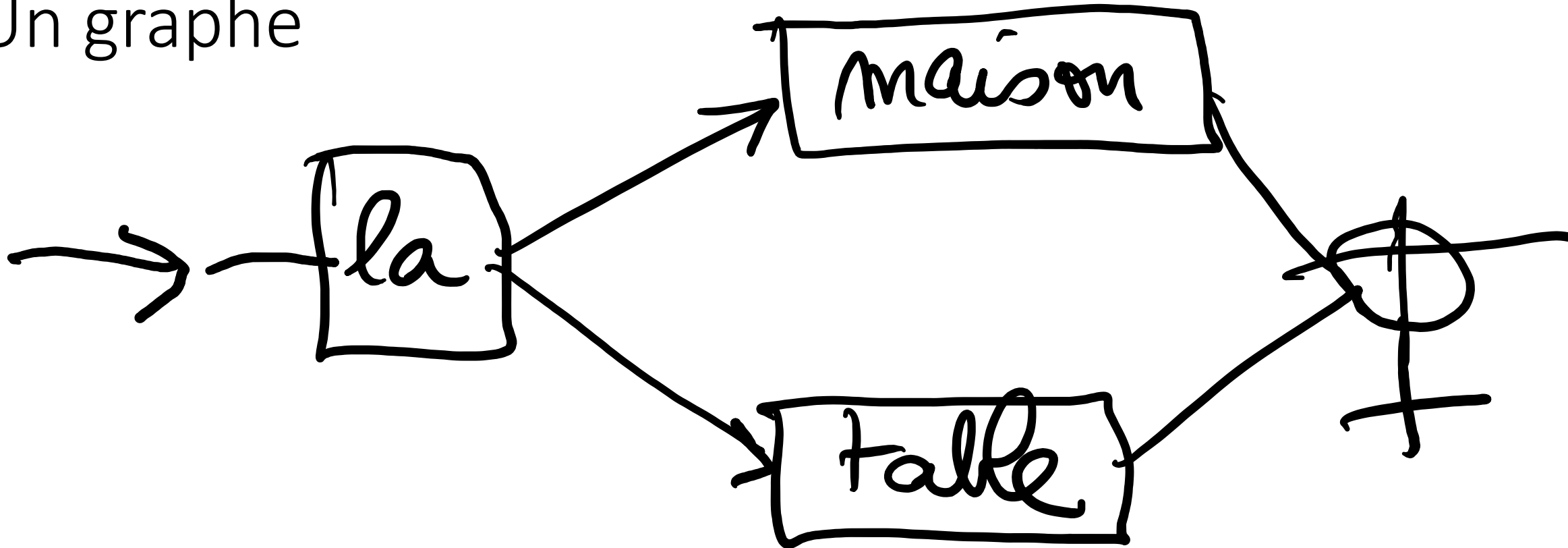


5. Grammaires locales

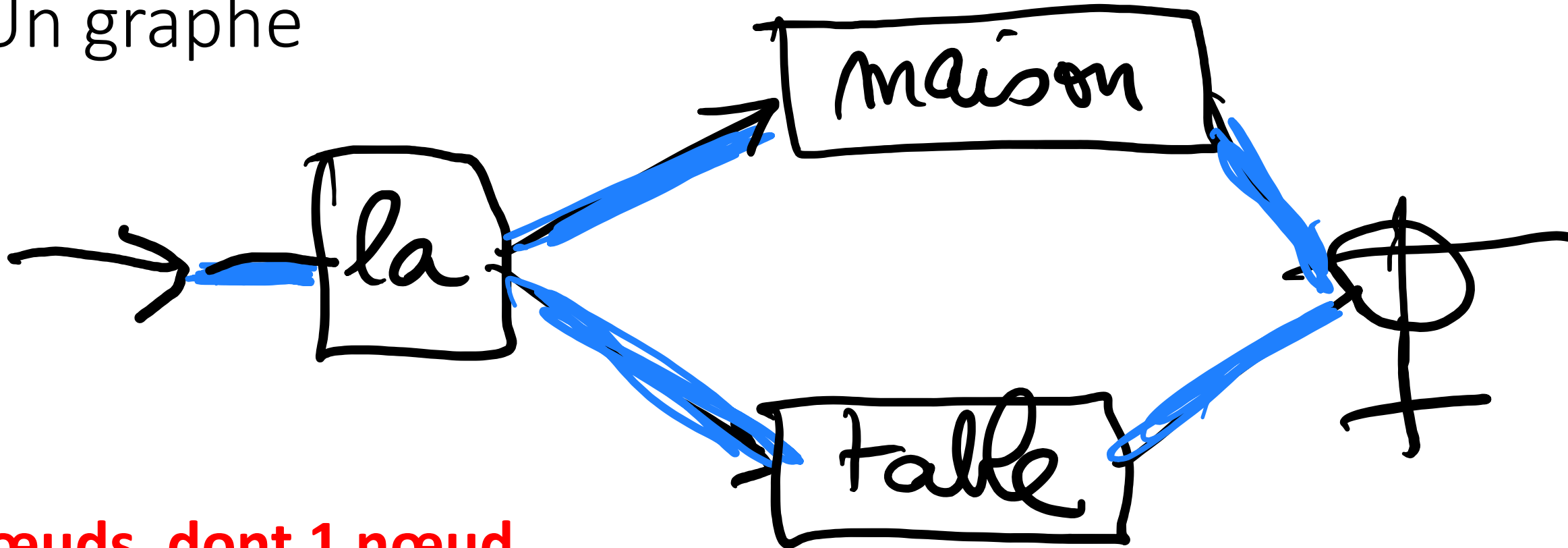
Graphes

- Un graphe est un ensemble de nœuds
- Un graphe comporte un nœud initial, et un nœud terminal
- Les nœuds sont étiquetés par des expressions
- Les nœuds peuvent être connectés ou non ; les connections sont orientées

Un graphe



Un graphe



5 nœuds, dont 1 nœud initial et 1 nœud terminal

5 connections

Editeur de graphe

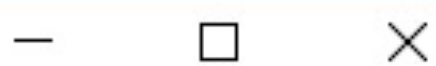
- CTRL – Cliquer pour créer un nouveau nœud et l'éditer
- CTRL – Entrée pour valider l'édition
- Cliquer sur le nœud source, **puis** sur le nœud cible pour les connecter
- Ne pas double-cliquer !

- CTRL – Z pour annuler la dernière opération ; CTRL – Y pour la ré-effectuer

L'éditeur de graphe

- Créer un graphe pour : *la (maison/table)*
- Ajouter *petite, une*

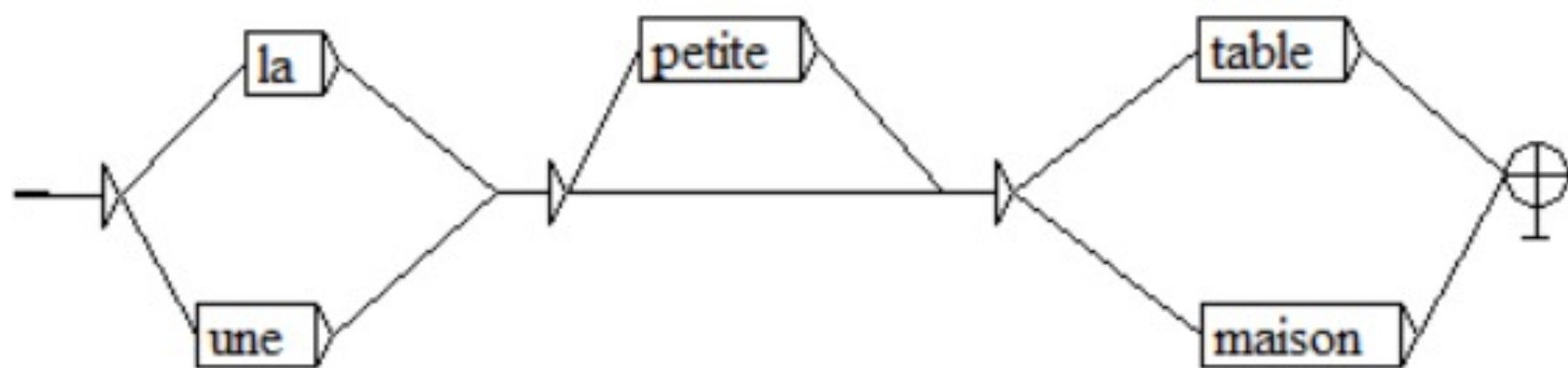
NooJ - [C:\Users\Max\Documents\NooJ\fr\Syntactic Analysis\zozo1423.nog]



File Edit Lab Project Windows Info GRAMMAR



French/French syntactic grammar.



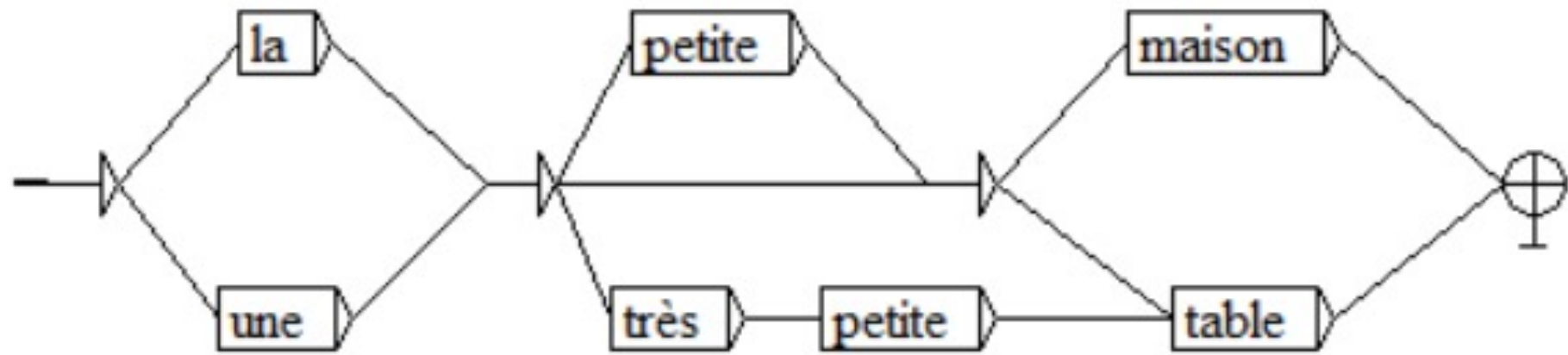
8.7 sec

Cancel

L'éditeur de graphe

- Créer un graphe pour : *la (maison/table)*
- Ajouter *petite, une*
- On veut en plus *très* dans "la très petite table" mais on ne veut pas "la très petite maison"

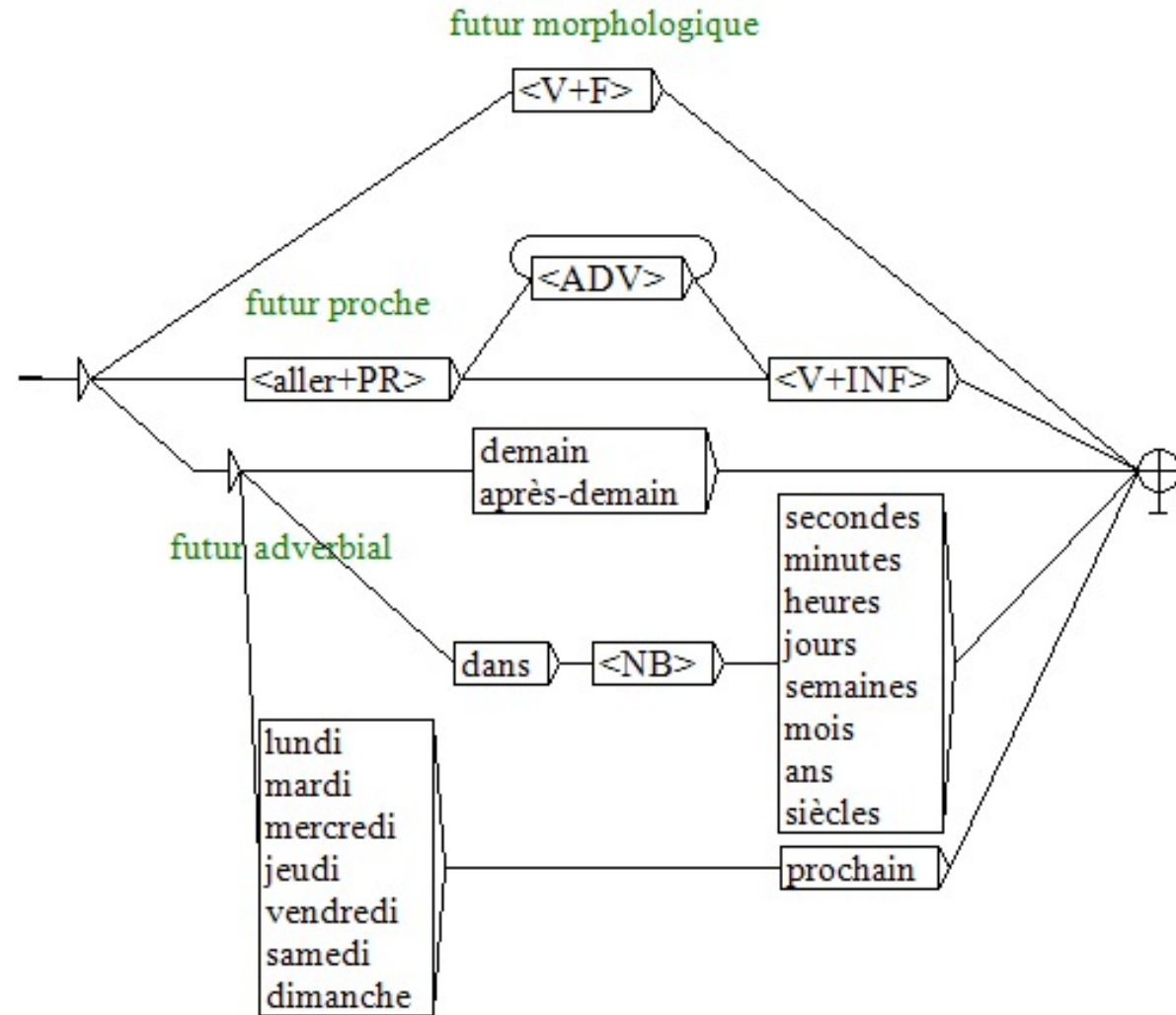
French/French syntactic grammar.



8.7 sec

Cancel

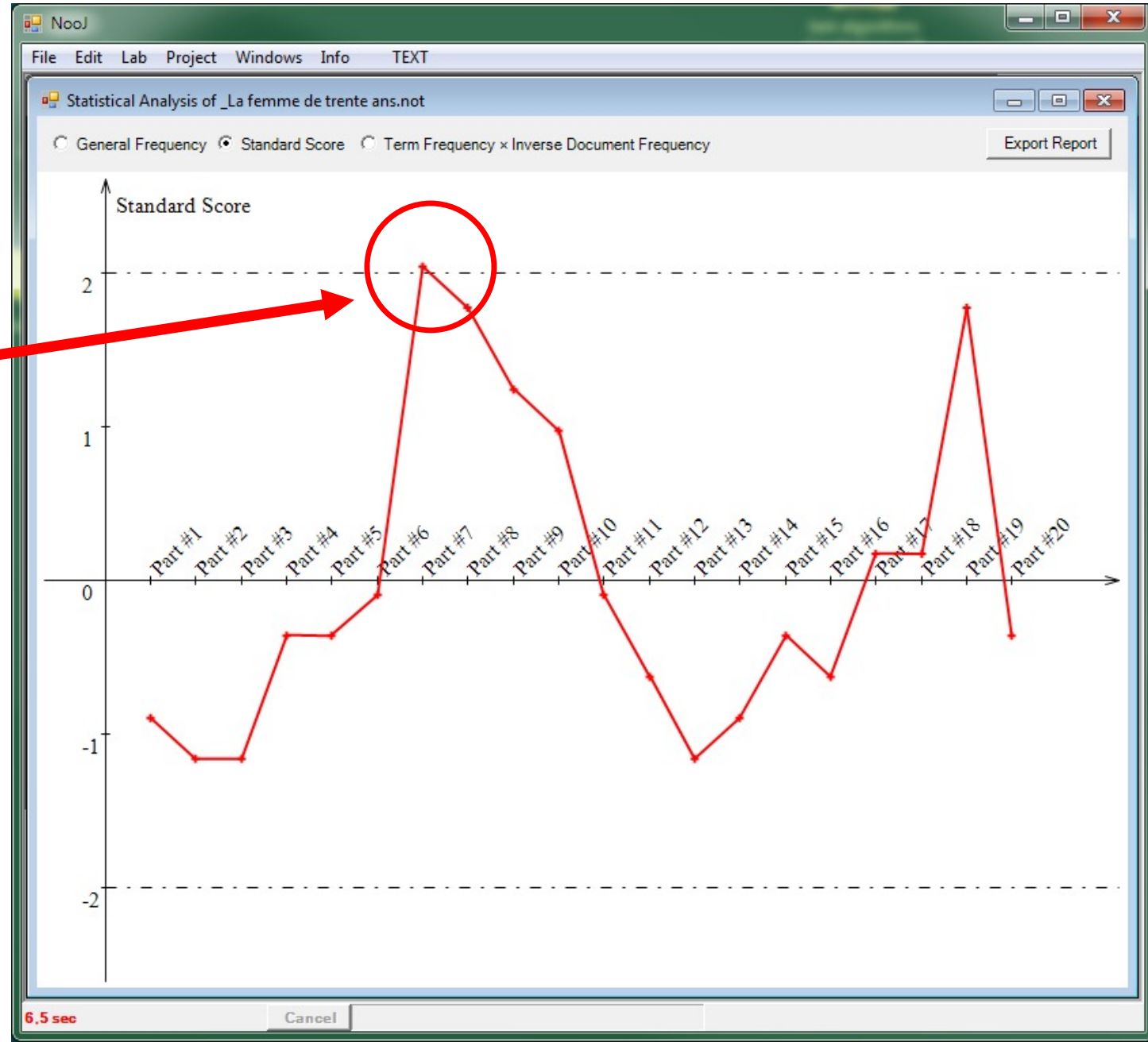
French (France)/French (France) syntactic grammar.



Analyse statistique d'une concordance : score standard

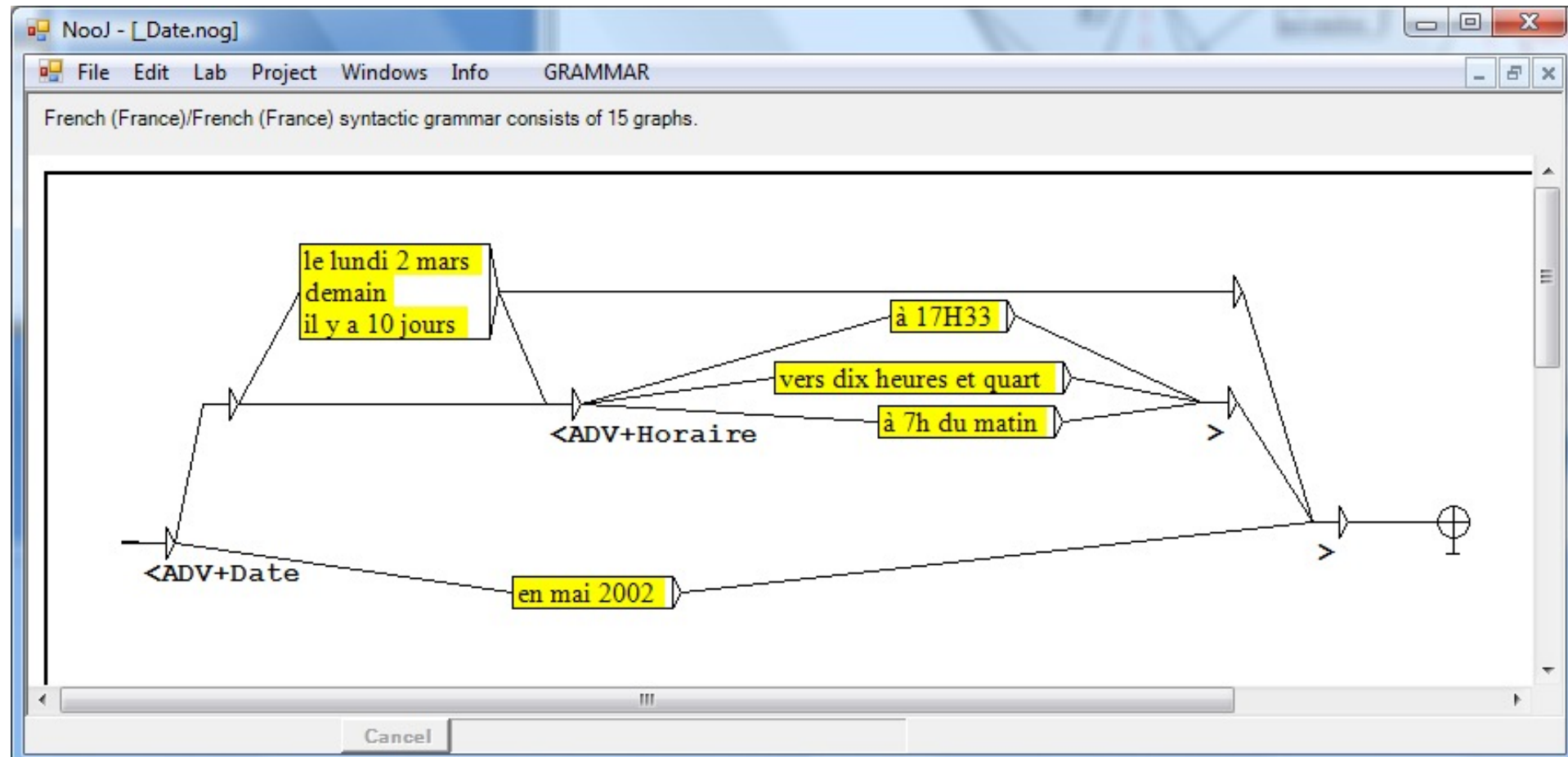
Intéressant

Le futur n'a pas été utilisé de façon anodine : il a été utilisé de façon anormalement élevée dans la partie 7 du texte : juste avant que le personnage principal ne se marie

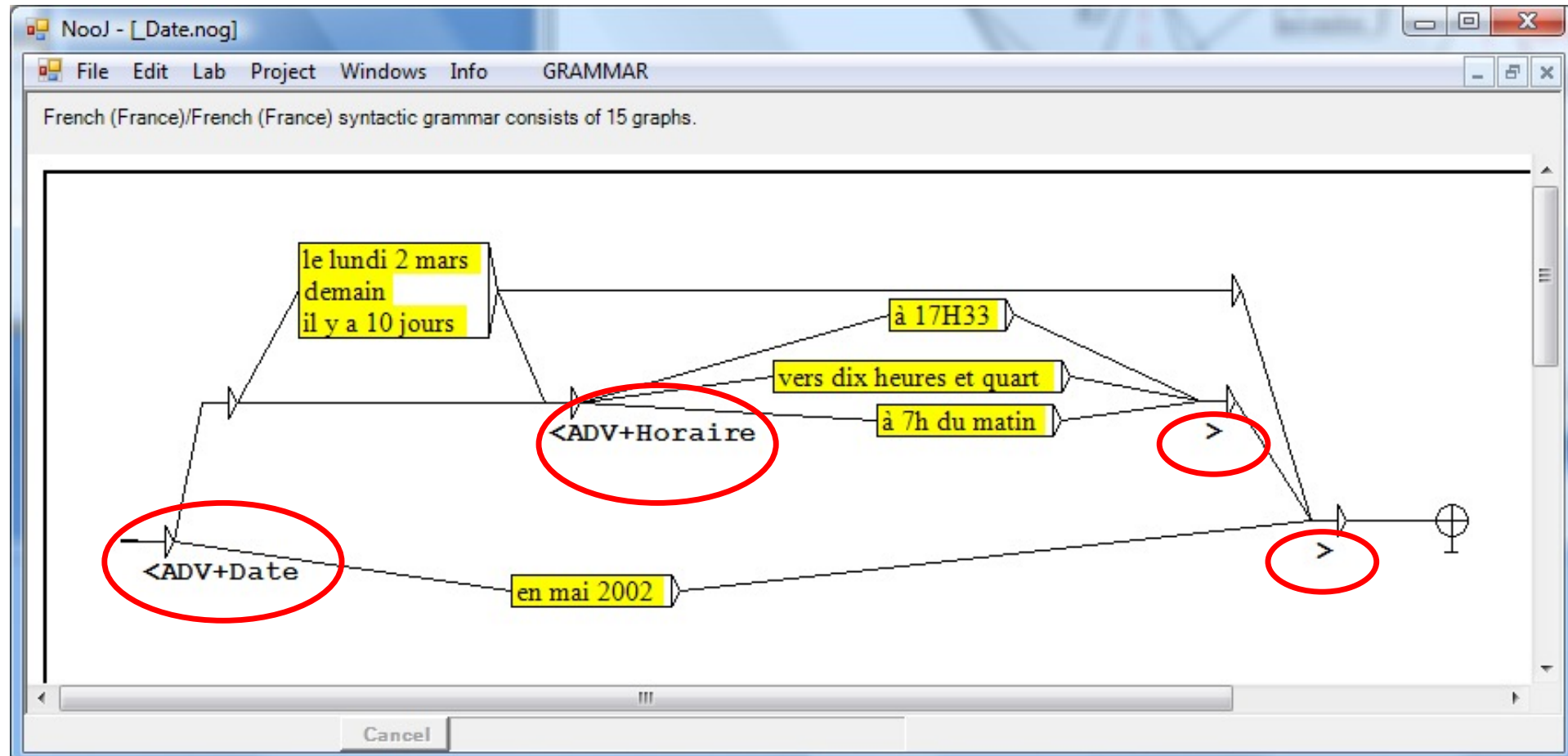


6. Annotation automatique

Annoter des textes



Annoter des textes



Annoter des textes

The screenshot shows the NooJ software interface with a text document titled "La femme de trente ans.not". The main window displays a paragraph of French text. A red box highlights the first token "1" in the annotation table, and a green oval highlights the label "ADV+Date" in the same row. A green arrow points from this label to the right, indicating its span across the text.

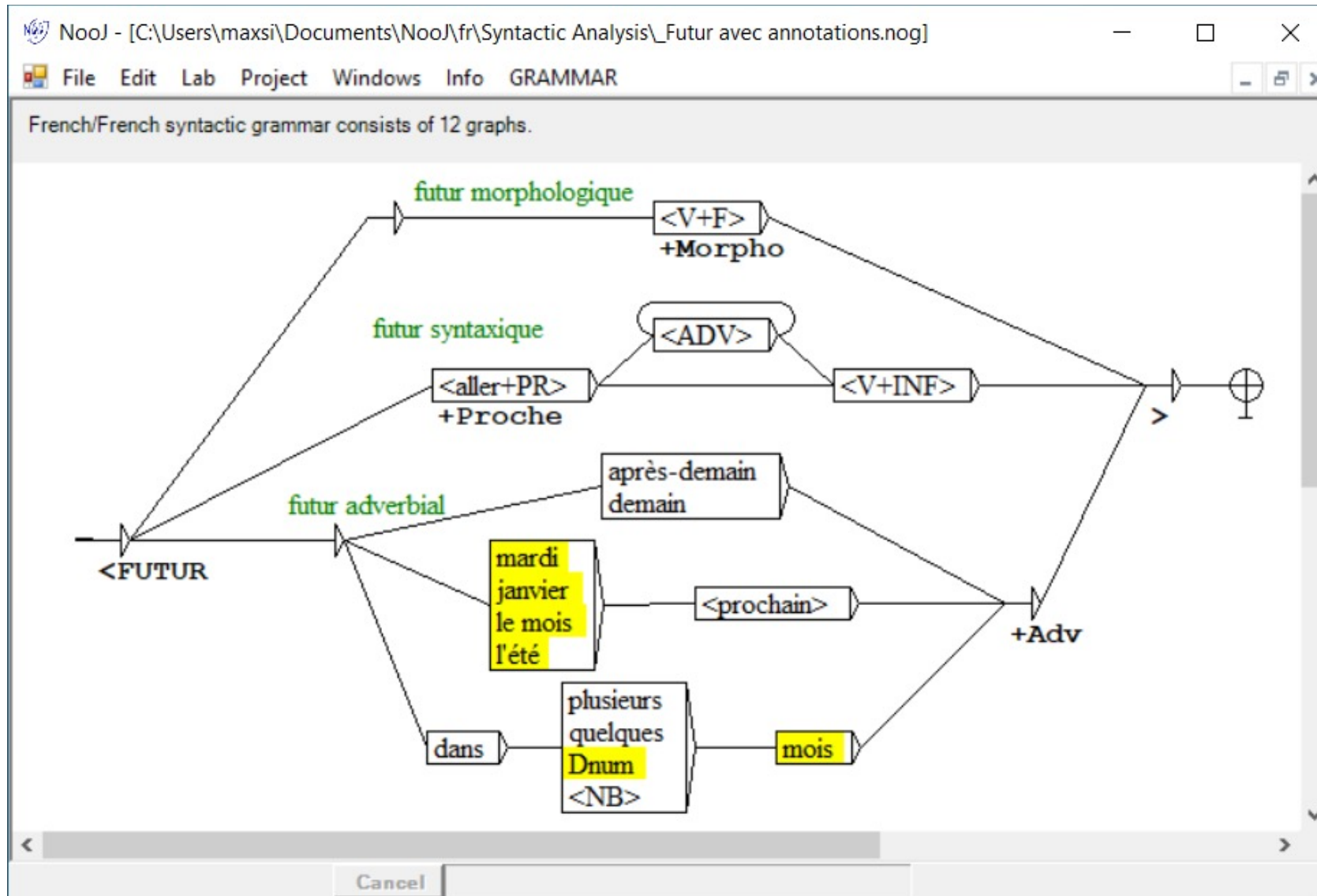
Language is "French (France) (fr)".
Text Delimiter is: \n (NEWLINE)
Text contains 926 Text Units (TUs).
81359 tokens including:
50000 word forms

Au commencement du mois d'avril 1813, il y eut un dimanche dont la matinée promettait un de ces beaux jours où les Parisiens voient pour la première fois de l'année leurs pavés sans boue et leur ciel sans nuages. Avant midi, un cabriolet à pompe attelé de deux chevaux fringants déboucha dans la rue de Rivoli par la rue Castiglione, et s'arrêta derrière plusieurs équipages stationnés à la grille nouvellement ouverte au milieu de la terrasse des Feuillants. Cette leste voiture était conduite par un homme en apparence soucieux et maladif, des cheveux grisonnants couvraient à peine son crâne jaune et le faisaient vieux avant le temps; il jeta les rênes au laquais à cheval qui suivait sa voiture, et descendit pour prendre dans ses bras une jeune fille dont la beauté mignonne attira l'attention des oisifs en promenade sur la terrasse. La petite personne se laissa complaisamment saisir par la taille quand elle fut debout sur le bord de la voiture, et passa ses bras autour du cou de son guide, qui la posa sur le trottoir, sans avoir chiffonné la garniture de sa robe en reps vert. Un amant n'aurait pas eu tant de soin. L'inconnu devait être le père de cette enfant qui, sans le remercier, lui prit familièrement le bras et l'entraîna brusquement dans le jardin. Le vieux père remarqua les regards émerveillés de quelques jeunes gens, et la tristesse empreinte sur son visage s'effaça pour un moment. Quoiqu'il fût arrivé depuis longtemps à l'âge où les hommes doivent se contenter des trompeuses jouissances que donne la vanité, il se mit à sourire:

1	1.01	4	17	17.01	20	25	25.01	27
ADV+Date								
à,PREP	le,DET+m+s	commencement,N+z1+m+s	du,DET+z1+m+s		mois d'avril,N+NDN+m+s			
			de,PREP	le,DET+m+s	mois,N+z1+m+p	de,PREP		avril,N+z1+m+s
					mois,N+z1+m+s	de,DET		

Cancel

Annoter des textes



C:\Users\maxsi\Documents\NooJ\fr\Projects_La femme de ...

- | 1 | + / 926 TUs

Characters Tokens Digrams

Language is Text Delimi: Text contai: 84963 token

Show Text Annotation Structure

Au commencement du mois d'avril 1813, il y eut un dimanche dont la matinée promettait un de ces beaux jours où les Parisiens voient pour la première fois de l'année leurs pavés sans boue et leur ciel sans nuages. Avant midi, un cabriolet à pompe attelé de deux chevaux fringants déboucha dans la rue de Rivoli par la rue Castiglione, et s'arrêta derrière plusieurs équipages stationnés à la grille nouvellement

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:

C:\Users\maxsi\Documents\NooJ\fr\Syntactic Analysis_Futur avec annotations.nog Set

Syntactic Analysis

Index

Shortest matches

Longest matches

All matches


Limitation

All occurrences

Only: 100 occ.

1 occ. per match

Reset Concordance



Concordance

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

Text	Before	Seq.	After
	est heureuse aujourd'hui, le bourdonnant comme un essaim. 'Cela	sera/<FUTUR+Morpho>	-t-elle toujours ?' Car le
	du Carrousel des régiments qui	sera/<FUTUR+Morpho>	donc bien beau ? dem
	il a bon coeur, l'	vont manoeuvrer/<FUTUR+Proche>	. - Non, mon enfant, to
	fait colonel... - Ma chère, Victor	entraînera/<FUTUR+Morpho>	peut-être à donner sa
	moi. Comment espérer que vous	restera/<FUTUR+Morpho>	colonel toute sa vie. Je
		nourrez/<FUTUR+Morpho>	vous entendre tous de

Query 244/244

NooJ

File Edit Lab Project Windows Info TEXT

[Modified] C:\Users\maxsi\Documents\NooJ\fr\Projects_La femme de trente ans.not

- 27 + / 926 TUs

Characters
Tokens
Digrams

Language is "French(fr)".
Text Delimiter is: \n (NEWLINE)
Text contains 926 Text Units (TUs).
84963 tokens including:
69739 word forms
62 digits

Show Text Annotation Structure

"Mais, mon père, disait-elle, il y a encore sur la place du Carrousel des régiments qui vont manoeuvrer.

- Non, mon enfant, toutes les troupes défilent.

88

93

spec=int → FUTUR+Proche →

spec=rel+Fonc=nom → aller, V+Aux=e+Nb=p+Pers=3+Temps=PR → manoeuvrer, V+Aux=a+Temps=INF →

Concordance

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

Text	Before	Seq.	After
	est heureuse aujourd'hui, le	sera/<FUTUR+Morpho>	-t-elle toujours ?' Car le
	bourdonnant comme un essaim. 'Cela	sera/<FUTUR+Morpho>	donc bien beau ? dem.
	du Carrousel des régiments qui	vont manoeuvrer/<FUTUR+Proche>	. - Non, mon enfant, to
	il a bon coeur, l'	entraînera/<FUTUR+Morpho>	peut-être à donner sa
	fait colonel... - Ma chère, Victor	restera/<FUTUR+Morpho>	colonel toute sa vie. Je
	moi. Comment espérer que vous	nourrez/<FUTUR+Morpho>	vous entendre tous de

Query 244/244

[Modified] C:\Users\maxs\Documents\NooJ\fr\Projects_La femme de tre...
 - | 27 + / 926 TUs
 Characters Tokens Digrams
 Language is "French(fr)"
 Text Delimiter is: \n
 Text contains 926 Text
 84963 tokens including:
 8000 word forms

Show Text Annotation Structure

"Mais, mon père, disait-elle, il y a encore sur la place du Carrousel des régiments qui vont manoeuvrer.

- Non, mon enfant, toutes les troupes défilent.

- Je pense, mon père, que vous vous trompez. M. d'Aiglemont a dû les faire avancer...

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:

<DET> <N> <FUTUR>

Syntactic Analysis

Index

- Shortest matches
- Longest matches
- All matches

Limitation

- All occurrences
- Only: 100 occ.
- 1 occ. per match

Reset Concordance

N O J

Concordance

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

word forms

Text	Before	Seq.	After
	des volontés différentes dont les	tyrannies seront	inconciliables ? Tu seras ou vict
	dit, pensa-t-elle, mon	neveu subira	bientôt les inconvénients du ma
	un de tes plus poignants	regrets viendra	du souvenir de ce que
	nous peignaient si délicieuses. Cette	soirée fera	ton désespoir, Louisa. Dans ce
	salle; quand il ne nous	y verra	plus, le jeune gentilhomme disc
	v verra plus. le ieune	gentilhomme discontinuera	de vous aimer par la

Query 31/31

TEXT > Export Annotations As XML

The screenshot illustrates the process of exporting annotations from NooJ. The main window shows a concordance table with the following content:

Text	Before	Seq.
Pierre	viendra/	<FUTUR+Mo
Pierre viendra	mardi prochain/	<FUTU
Pierre viendra mardi prochain ou	dans quinze jours/	<FU

The 'Export date' dialog box is open, showing the following options:

- Tag all syntactic annotations (selected)
- Tag only following annotations (click arrow for examples)
- Export Annotated Text Only (ignore non-annotated text)

The success message dialog box indicates: "Success: text has been exported into file C:\Users\Max\Documents\NooJ\fr\Projects\date.not.xml.txt".

The Notepad window shows the resulting XML output:

```
Pierre <FUTUR TYPE="Morpho">viendra</FUTUR>
<FUTUR TYPE="Adverbial">mardi prochain</FUTUR>
ou <FUTUR TYPE="Adverbial">dans quinze jours</FUTUR>.
```


Automatic HTML Annotation (Legifrance database)

L'acte d'engagement, qui constitue au sens de l'[article 11 du code des marchés publics](#) l'offre des candidats, est nécessairement un document signé par les parties et notamment par le cocontractant de l'administration. D'ailleurs, en cas de contestation, sera qualifié comme tel le document signé par les parties quand bien même l'administration contractante ne lui aurait pas réservé cette appellation ou qu'elle l'ait attribuée à un document autre parmi ceux qui composent les pièces du marché (cf. [CE 24 avril 1992, Syndicat mixte pour la géothermie à La Courneuve, req. no 112679](#), Lebon T., p. 1124, 1263 et 1264, voir [II.201.1/1](#)). L'acte d'engagement, qui constitue au sens de l'[article 11 du code des marchés publics](#) l'offre des candidats, est nécessairement un document signé par les parties et notamment par le cocontractant de l'administration. D'ailleurs, en cas de contestation, sera qualifié comme tel le document signé par les parties quand bien même l'administration contractante ne lui aurait pas réservé cette appellation ou qu'elle l'ait attribuée à un document autre parmi ceux qui composent les pièces du marché (cf. [CE 24 avril 1992, Syndicat mixte pour la géothermie à La Courneuve, req. no 112679](#), Lebon T., p. 1124, 1263 et 1264, voir [II.201.1/1](#)).

Enfin, il convient de rappeler que le défaut de signature de l'acte d'engagement par le cocontractant de l'administration constitue une irrégularité substantielle (cf. [CE 10 décembre 1993, Société Lopez Entreprise, req. no 124529](#), Mon. TP, 7 janvier 1994, p. 41 voir [II.241.1/1](#)) de même que l'omission sur ce même document de la mention de l'identité du cocontractant (cf. [CE 14 janvier 1998, Commune de Saint-Pierre \(Réunion\) c/ M. Pihouée, req. no 154929](#), voir [II.201.2/1](#)). Enfin, il convient de rappeler que le défaut de signature de l'acte d'engagement par le cocontractant de l'administration constitue une irrégularité substantielle (cf. [CE 10 décembre 1993, Société Lopez Entreprise, req. no 124529](#), Mon. TP, 7 janvier 1994, p. 41 voir [II.241.1/1](#)) de même que l'omission sur ce même document de la mention de l'identité du cocontractant (cf. [CE 14 janvier 1998, Commune de Saint-Pierre \(Réunion\) c/ M. Pihouée, req. no 154929](#), voir [II.201.2/1](#)).

Accepté par le cocontractant, l'acte d'engagement est signé par l'autorité administrative qualifiée pour conclure le contrat en vertu des attributions législatives ou réglementaires de compétence. L'[article 20 du code des marchés publics](#) qualifie les personnes ayant reçu compétence pour signer les marchés de « *personne responsable du marché* » (par abréviation : la PRM), l'expression couvrant les autorités tant étatiques que locales. Accepté par le cocontractant, l'acte d'engagement est signé par l'autorité administrative qualifiée pour conclure le contrat en vertu des attributions législatives ou réglementaires de compétence. L'[article 20 du code des marchés publics](#) qualifie les personnes ayant reçu compétence pour signer les marchés de « *personne responsable du marché* » (par abréviation : la PRM), l'expression couvrant les autorités tant étatiques que locales.

Les règles de compétence sont d'ordre public et leur méconnaissance entache le marché d'illégalité. Il importe donc de veiller à ce qu'elles soient strictement respectées. Les règles de compétence sont d'ordre public et leur méconnaissance entache le marché d'illégalité. Il importe donc de veiller à ce qu'elles soient strictement respectées.

Les autorités administratives qualifiées pour signer l'acte d'engagement diffèrent selon que les marchés sont passés au nom de l'État, des collectivités territoriales locales ou des établissements publics (voir [figure 3](#)). Les autorités administratives qualifiées pour signer l'acte d'engagement diffèrent selon que les marchés sont passés au nom de l'État, des collectivités territoriales locales ou des établissements publics (voir [figure 3](#)).

Les ministres, chacun pour son département ministériel, constituent les autorités légalement qualifiées pour conclure les marchés passés au nom de l'État. Les ministres signent l'acte d'engagement du marché à moins qu'ils ne délèguent leur compétence ou leur signature à des fonctionnaires de leur département ministériel. De part les dispositions de l'[article 20 du code des marchés publics](#), il leur est en effet loisible : Les ministres, chacun pour son département ministériel, constituent les autorités légalement qualifiées pour conclure les marchés passés au nom de l'État. Les ministres signent l'acte d'engagement du marché à moins qu'ils ne délèguent leur compétence ou leur signature à des fonctionnaires de leur département ministériel. De part les dispositions de l'[article 20 du code des marchés publics](#), il leur est en effet loisible :

- de déléguer leur compétence ou leur signature aux personnes responsables des marchés (PRM) ; de déléguer leur compétence ou leur signature aux personnes responsables des marchés (PRM) ;
- de réserver à leur propre signature certaines catégories de marchés. de réserver à leur propre signature certaines catégories de marchés.

La signature des marchés passés au nom de l'État, pour le compte des services déconcentrés des administrations civiles, ressortit à la compétence des préfets, eu égard à leur qualité de représentants de l'État dans les départements et les régions (D. no 82-389 et no 82-390 du 10 mai 1982 interprétés par une circulaire du Premier ministre en date du 12 juillet 1982. – Cf. [CAA Nancy 28 avril 1992, Société Quille et société Bouygue, req. no 91NC00098](#), Lebon T., p. 1111, voir [II.211.1/2](#)). Ces derniers tirent des textes qui ont aménagé l'exercice des compétences étatiques dans le cadre régional et départemental le droit de déléguer leur signature aux chefs des différents services déconcentrés. La signature

Des grammaires locales pour reconnaître des noms de technologie, de personnes/fonctions, d'institution et d'organisation, de dates, d'adresses, d'expressions de mesures, etc.

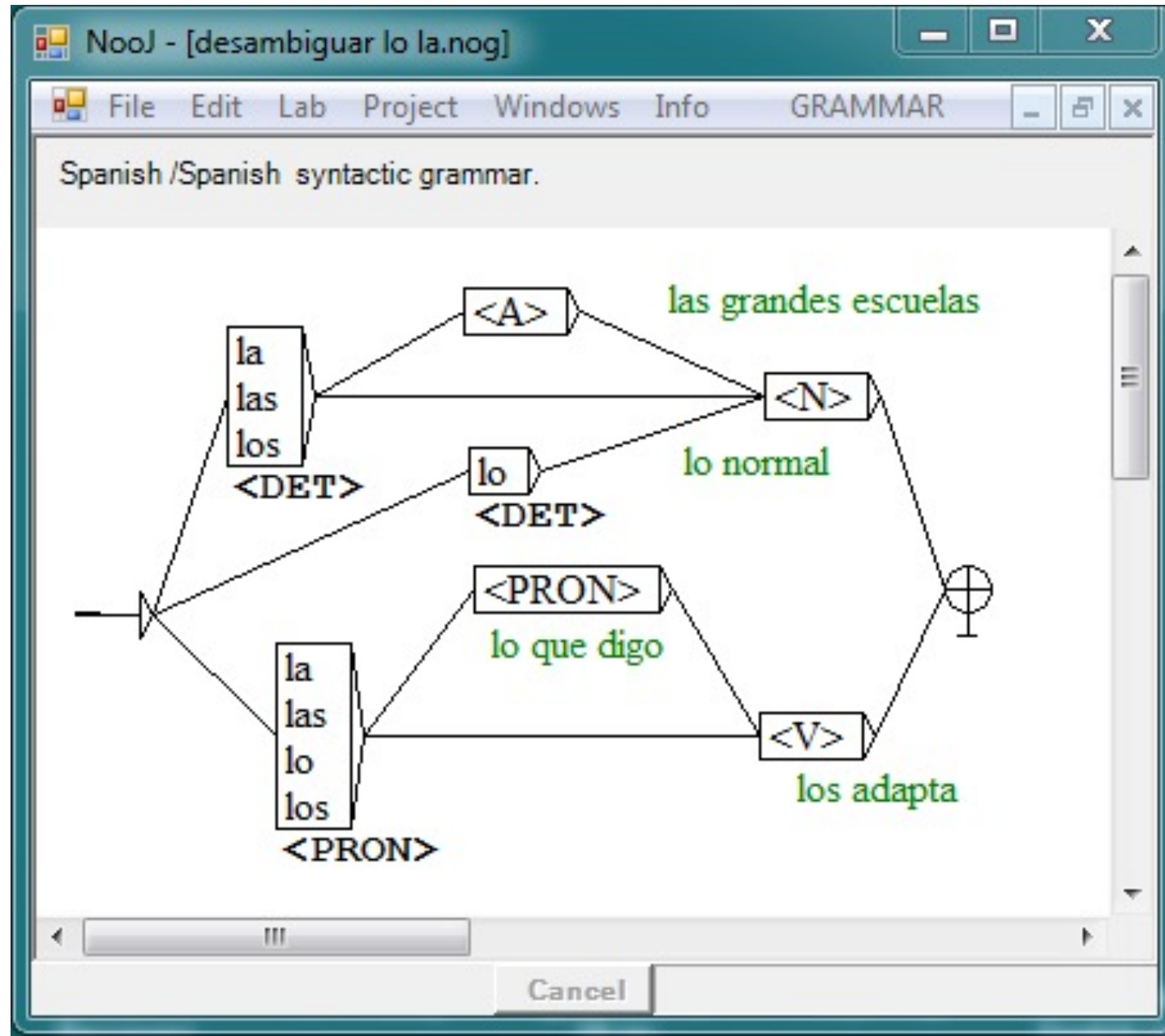
Concordance for Corpus Corpus corpusorg.noc

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

word forms

Text	After	Seq.
text 1.not	مدينة الزقازيق " محافظة الشرقية " 8- جامعة	جامعة الزقازيق
text 1.not	أسسيوط "الصعيد" 10- جامعة المنيا ----- المنيا	جامعة أسسيوط
text 1.not	المنصورة " محافظة الدقهلية " 13- جامعة قناة	جامعة المنصورة
text 1.not	طريق الاسماعيليه " الدراسة باللغة الانجليزية	جامعة مصر الدولية
text 1.not	مدينة 6 اكتوبر أولا : التخصصات العلمية	جامعة 6 اكتوبر
text 2.not	- أكدا ص.ب. 554 شاله أكدا	جامعة محمد الخامس
text 2.not	ص.ب. 19-9167 زنقة طارق بن	جامعة الحسن الثاني
text 2.not	ص.ب. 524 وجدة - المركب الجامعي	جامعة محمد الأول
text 2.not	للبنزول والمعادن ص.ب 31261 الظهران	جامعة الملك فهد
text 2.not	92 شارع 9 أفريل 1938- تونس- 1007 تليفون: 21671-562700/567322 فاكس	جامعة تونس
text 2.not	بقرطاج 4. (جامعة الحقوق الاقتصادية و	جامعة 7 نوفمبر
text 2.not	(الجنوب) طريق المطار كم 8.5 صفاقس	جامعة صفاقس
text 2.not	دمشق، البرامكة- الجمهورية العربية السورية	جامعة دمشق
text 2.not	حلب، الجمهورية العربية السورية تليفون	جامعة حلب
text 2.not	ص.ب (50) الخوض 123- سلطنة عمان	جامعة السلطان قابوس
text 2.not	بيروت، ص.ب 115020 بيروت تليفون	جامعة بيروت العربية
text 2.not	ص.ب-21- محافظة النجف - الكوفة	جامعة الكوفة
text 2.not	جامعة الموصل-العراق تليفون: 96460-810733 فاكس	جامعة الموصل
text 2.not	الجادرية - بغداد - العراق تليفون : 9641-7760735 فاكس	جامعة بغداد
text 2.not	ص ب 321 - الرمز البريدي 11115 تليفون	جامعة الخرطوم
text 2.not	تليفون : 0811)3-227986 (جوبا فاكس : (00249-11-222142) ص.ب	جامعة جوبا

Des grammaires locales pour lever des ambiguïtés



7. Bonus : levée d'ambiguïté sémantique

Les Verbes Français (Dubois & Dubois-Charlier) : 25.000 sens différents décrits soigneusement

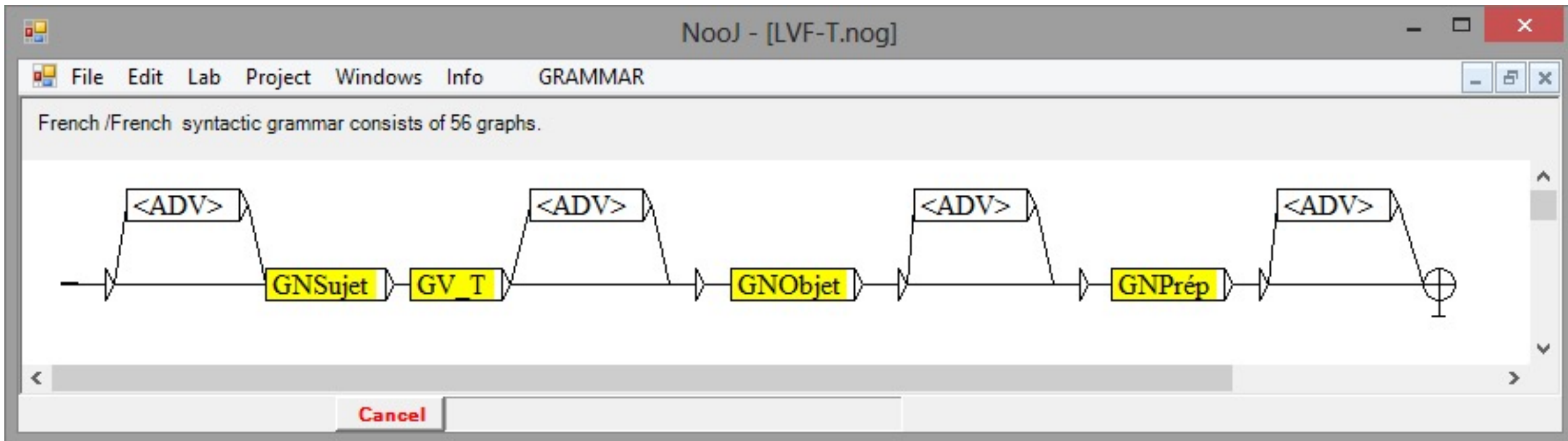
	A	B	C	D	E	F
1	M	DOM	CLASSE	OPER	SENS	PHRASE
2	abaisser 01	LOC	T3c	r/d bas qc	baissér	On a~le rideau de fer,le store.Le rideau du magasin s'a~.
3	abaisser 02	TEC	E3f	f.ire qc VRS bas	incliner,pencher	On a~la manette,le levier.La manette s'a~vers le bas.
4	abaisser 03	QUA	M3c	f.mvt mns hauteur	baissér	On a~le mur d'un mètre.Le mur s'a~de beaucoup.
5	abaisser 04	MON	M4b	f.mvt mns val	baissér	On a~les prix,les revenus de dix p.c.Les prix s'a~de bcp.
6	abaisser 05	MED	T4b	r/d bas quant	faire descendre	Le malade a~la fièvre avec l'aspirine.La fièvre s'a~.
7	abaisser 06	PSYt	R2a	m.e.état mvs qn abs	avilir,humilier	On a~P,son orgueil en le blâmant.On s'a~en public.
8	abaisser 07(s)	PSY	M2b	f.mvt humble A+inf	s'humilier à	On s'a~à demander une faveur,à cette demande.
9	abaisser 08(s)	VEH	L3a	(voie)lc.qp vrs bas	descendre vers	La route s'a~vers la rivière.
10	abaisser 09(s)	PSY	M2b	f.mvt A état vil	s'avilir,tomber jusqu'à	On s'a~au niveau de cet escroc.
11	abalourdir (s)	PSYt	T2b	r/d balourd-a	abêtir,abrutir	On s'a~avec un tel film.On a~P avec ce livre.La télé a~P.
12	abandonner 01	DRO	D2a	dat qc A qn	laisser,léguer	On a~ses biens à ses enfants,à une fondation.
13	abandonner 02	MAR	D3a	dat mvt A qc	laisser aller	On a~sa barque à un fort courant.
14	abandonner 03	EQU	S3b	dgrp rênes	lâcher,laisser aller	On a~les rênes de l'attelage.
15	abandonner 04	PSY	S4b	dgrp abs	renoncer à	On a~un projet,une idée.Le projet est a~.
16	abandonner 05	SOC	S4b	dgrp abs	quitter,lâcher	On a~ses études.On n'a~pas après un premier échec.
17	abandonner 06	SPO	S2e	dgrp lutte	s'avouer vaincu	On a~le match.Le boxeur a~au premier round.
18	abandonner 07	COM	T3f	li.mut CT/PR	quitter	On a~un appartement pour une maison.
19	abandonner 08	SOC	S3h	dgrp lc	quitter,inhabité	On a~ce village devenu désert.Le village est a~.
20	abandonner 09	LOC	S2d	dgrp qn qp	quitter,laisser	On a~un chien sur la route,un enfant dans la voiture.
21	abandonner 10	SOC	S4b	dgrp abs	renoncer à	On a~ses responsabilités.Après cet échec,on a~.
22	abandonner 11	SOC	S2d	(qc)dgrp qn	quitter,manquer à	Ses forces,son énergie,la chance ont a~P.
23	abandonner 12(s)	PSY	P1h	ger.mens A qc/inf	se laisser aller à	On s'a~à un grand désespoir,à désespérer.
24	abandonner 13(s)	LANt	C1a	loq confidence	s'épancher	On s'a~dans la détresse.
25	abasourdir 01	PAT	H1b	f.som étourdi	étourdir,assourdir	On a~P avec la radio.Le bruit a~P.
26	abasourdir 02	PSY	P2a	(abs)f.sent vif à qn	stupéfier	La nouvelle a~P.On reste a~après cette annonce.
27	abâtardir 01	LITt	T4b	r/d bâtard abs	dégénérer,s'altérer	On a~la littérature avec ces publications.Le roman s'a~.
28	abâtardir 02(s)	BIOt	T3a	(qc)d bâtard	dégénérer	La race s'a~,est a~.Un métissage a~cette race.
29	abattra 01	OBJ	E3f	f.ire qc e bas	renverser,f tomber	On a~les arbres sur le sol.Le chêne est a~par la tempête.
30	abattra 02	MII	F3c	ict mvs sérn	descendre	On a~un avion avec un missile.

Lexique + Morphologie + Syntaxe + Sémantique

- *abriter* #1 (T11b8, P10b8) : *Luc abrite Léa de la pluie avec un parapluie*
[Luc shelters Lea from the rain with an umbrella]
- *abriter* #2 (T1101, P1001) : *Luc abrite des réfugiés chez lui*
[Luc gives shelter to refugees]
- *abriter* #3 (T3100) : *Cet immeuble abrite le service juridique*
[This building houses the legal department]
- *abriter* #4 (P10b1) : *Luc s'abrite des ennuis derrière un prétexte*
[Luc hides behind a pretext]
- *abriter* #5 (T13b8) : *On abrite le port des vagues avec des digues*
[One shields the harbour from the waves with a seawall]

Grammaire générique Txxxx

(phrases transitives directes avec second complément indirect)



Grammaire T3100 → sens #3

centre de Pristina	, il abritait plusieurs partis politiques des minorités ethniques	. L'attentat visait
selon vous, Monsieur	, le bâtiment qui abrite la plus grande quantité de rêves	? L'école? Le
chimiques et bactériologiq...	Ce pays abrite le plus grand réseau mondial de drogue	, et ses banques
froideur presque menaçante	, il abrite des banques	et un inévitable
de produits nucléaires...	Chaque pays abrite ses milieux criminels	. Les principales orga
jeûne pour 5 livres (6)	: elle abrite un cirque	(sans animaux), un
Alexandra et Lenasia	, elle abrite deux millions cinq cent mille personnes	. Un grand bidonville
'Etat de droit.	Le Cambodge abrite une société structurée	non pas en
ses contradictions internes.	Ce pays montagneux abrite un grand nombre de groupes ethniques	, y compris parmi
terre de Flandre	. la ville qui abrite la Commission européenne	suscite bien des

Grammaire P10b1 → sens #4

première fois aussi	, le premier ministre israélien ne s'abrita pas derrière les slogans traditionnels	- "Jérusalem ui
désarmé. En Europe	, il s'abrite derrière les pressions exercées	par le Luxemb
la corruption endémique.	Les responsables politiques s'abritent derrière le carcan traditionnel	des "valeurs a
prend pas parti.	Il s'abrite derrière sa tâche d'homme	de théâtre qui
dissimuler le fond	, il s'abrite en général derrière des considérations de forme,	et remet tout
système financier international.	Il s'abrite derrière le contentieux colonial,	dont il exige

Conclusion

- On peut construire des applications de recherche d'information, d'annotation automatique et d'extraction d'information en appliquant des ressources linguistiques construites soigneusement ;
- Les ressources linguistiques ont la forme de dictionnaires (listes de taille finie) ou de grammaires (potentiellement infinies)
- La plateforme NooJ peut être utilisée pour développer ces ressources, les accumuler, les tester et les corriger, les partager et assurer leur maintenance
- En traitant des unités décrites linguistiquement, on obtiendra des résultats d'analyse bien plus fiables qu'en traitant des formes graphiques.

www.nooj4nlp.org