

NooJ

Un outil linguistique au service des humanités numériques

<https://nooj.univ-fcomte.fr>

max.silberztein@univ-fcomte.fr

Outils informatiques pour analyser des corpus de textes : deux approches

Les outils informatiques utilisés pour analyser des corpus de textes écrits en langue naturelle suivent deux approches :

- L'approche **formelle** : on décrit précisément les expressions en jeu, en utilisant des dictionnaires et des grammaires construites par des spécialistes
- L'approche **empirique** : on utilise des corpus d'apprentissage pré-analysés que l'on compare avec le texte traité, et on copie les analyses préconstruites.

Succès de l'approche empirique pour l'intelligence artificielle

AlphaGo versus Lee Sedol

[Article](#) [Talk](#)

From Wikipedia, the free encyclopedia

AlphaGo versus Lee Sedol, also known as the **DeepMind Challenge Match**, was a five-game [Go](#) match between top Go player [Lee Sedol](#) and [AlphaGo](#), a [computer Go](#) program developed by [DeepMind](#), played in [Seoul](#), South Korea between 9 and 15 March 2016. AlphaGo won all but the fourth game;^[1] all games were won by resignation.^[2] The match has been compared with the historic chess match between [Deep Blue and Garry Kasparov](#) in 1997.

ChatGPT and the spectacular evolution of natural language processing

25 APR. 2023

The buzz around ChatGPT raised interest on recent advances in natural language processing based on pre-trained artificial intelligence models and on the principle of transfer learning. Michalis Vazirgiannis and Moussa Kamal Eddine are specialists of this field in the Data Science and Mining group at the Computer Science Laboratory of the Ecole Polytechnique. They explain how ChatGPT work, the latest developments and the issues at stake.

L'intelligence artificielle **spectaculaire** en 2024

- ChatGPT, Microsoft Copilot, Google Gemini
- Gamma (produit des présentations Powerpoint)
- Copy.AI, Jasper, Madgicx (produit des campagnes de marketing)
- Midjourney, Dall-E3 (produit des images à parti de leur description)
- Lex (produit des réseaux d'idées « brainstorming »)
- Speechify (Text to Speech)
- Grammarly Business (corrige fautes d'orthographe et style)
- Fathom (gère les idées produites lors de meetings)
- Elicit (produit des résumés d'articles scientifiques)
- Perplexity AI (moteur de recherche intelligent)
- Sybill (produit des emails de relance)
- Tesla (voitures autonomes)
- ...

C'est magique !



Siri. Quelqu'un à qui parler.

Siri est l'assistant intelligent qui comprend ce que vous dites, voit ce que vous voulez dire et trouve les réponses à vos questions. Et comme il maîtrise désormais davantage de langues, il se rend indispensable dans plus de pays².

Promesses, promises



Wired

<https://www.wired.com> > ... > Electric Vehicles

Elon Musk Says Every New Tesla Can Drive Itself

Oct 19, 2016 — Tonight, at a press conference, he announced that every new Tesla will be fully capable of driving itself. After being upgraded with a suite of ...



← Tweet



Elon Musk 
@elonmusk

In ~2 years, summon should work anywhere connected by land & not blocked by borders, eg you're in LA and the car is in NY

9:11 PM · Jan 10, 2016

4,208 Retweets 534 Quote Tweets 7,414 Likes



17:27

5G



electrek



Tesla's self-driving is still not working in Boring Company's one-way tunnels

Fred Lambert | Aug. 27th 2024 12:12 pm PT

Tesla's self-driving system is still not working in The Boring Company's one-way tunnels under Las Vegas despite reportedly working on it for years.

But self-driving in Tesla vehicles on public roads is around the corner?

Tesla uberbulls often like to say that Tesla is the leader in self-driving because while it doesn't have a commercially available autonomous ride-hailing service like Waymo, it doesn't rely on geo-fencing and manning like Waymo.



mais avec quelques bémols tout de même (hallucinations)



How many countries begin with the letter "Y"?



As of my last knowledge update in January 2022, there are three countries whose names start with the letter "Y." These countries are:

1. Yemen
2. Zambia
3. Zimbabwe

Please note that geopolitical changes can occur, and the number of countries in the world can change over time, so it's a good practice to verify this information with up-to-date sources for the most current and accurate count.



Lindsay McGregor

@LindsMcGregor · Jul 5

46 330

Me: Alexa, turn on family room lights.
Alexa: Okay, playing "Livin' on a Prayer, lullaby version."
#alexafail #robottakeover 🤪



© Twitter

Free-styling! Alexa decided she'd rather turn on a lullaby than the lights... surely making for a rather dark musical experience



If 1 woman can make 1 baby in 9 months, how many months does it take 9 women to make 1 baby? Explain each step you used to arrive at your answer.



First, we need to determine how many babies each woman makes per month. We can do this by dividing the number of babies each woman makes (1) by the number of months it takes for each woman to make a baby (9). This gives us a rate of $1/9$ babies per month for each woman.

Next, we need to determine how many babies 9 women make per month. We can do this by multiplying the number of women (9) by the rate at which each woman makes babies ($1/9$ babies per month). This gives us a total rate of $9 * (1/9) = 1$ baby per month.

Finally, we need to determine how many months it takes for 9 women to make 1 baby. We can do this by dividing the number of babies we want (1) by the rate at which the women are making babies (1 baby per month). This gives us a total of $1 / 1 = 1$ month.

Therefore, it takes 9 women 1 month to make 1 baby.


... et quelques problèmes éthiques

07:37 98

PC GAMER

OpenAI says it's 'impossible' to create ChatGPT without copyrighted content, as if that's somehow a good excuse

In the face of a growing number of lawsuits, OpenAI insists that the use of copyrighted content to train LLMs is fair use.



(Image Credit: Getty Images)

By Andy Chalk, published January 08, 2024

Just a couple weeks after being **sued by the New York Times** over allegations that it copied and used "millions" of copyrighted news articles to train its large-language models, OpenAI has told the UK's House of Lords communications and digital select committee (via **The Guardian**) that it has to

Today News+ Sports Audio Following

News+

THE NEW YORKER

More people are being arrested for crimes they have no connection to — and AI is to blame

Play Now 40 min

Today News+ Sports Audio Following

Approche empirique pour le Traitement Automatique des Langues

Automatic part of speech tagging is an area of natural language processing where statistical techniques have been more successful than rule-based methods. (Eric Brill, 1992)

The availability of large text corpora has changed the scientific approach to language in linguistics and cognitive science. (Manning, Schütze 2001)

Pas de linguistique à la conférence *Computational Linguistics* ???

THE 29TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS

TITLE
A Coarse-to-fine Cascaded Evidence-Distillation Neural Network for Explainable Fake News Detection
Finding Influential Instances for Distantly Supervised Relation Extraction
Stability of Syntactic Dialect Classification Over Space and Time
Hierarchical Attention Network for Explainable Depression Detection on Twitter Aided by Metaphor Concept Mappings
CMQA: A Dataset of Conditional Question Answering with Multiple-Span Answers
TSAM: A Two-Stream Attention Model for Causal Emotion Entailment
Few-shot Named Entity Recognition with Entity-level Prototypical Network Enhanced by Dispersedly Distributed Prototypes
Two Languages Are Better Than One: Bilingual Enhancement For Chinese Named Entity Recognition
DCT-Centered Temporal Relation Extraction
Debiasing Word Embeddings with Nonlinear Geometry
CofeNet: Context and Former-Label Enhanced Net for Complicated Quotation Extraction

Screenshot

Les applications de TAL

Aujourd'hui, le grand public, et les chercheurs en SHS (Histoire, Linguistique de corpus, Littérature, Socio-linguistique, Psychologie, etc.) disposent de nombreux outils pour traiter leur corpus, mais ces outils traitent des formes graphiques et non des unités linguistiques :

Alceste, Hyperbase, IRaMuTeQ, Lexico 3, Mallet, Sketch Engine, Le Trameur, TXM, Word2Vec, etc.

Chercher le concept *amour*

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

Cliquer sur un mot pour l'activer

← →

Index hiérarchique

Chercher un mot

Graphique

Éditer

Quitter

Retour

Sommaire

Formes	Lemmes	Chercher un code	Codes
111 amour	N° 1 TEX1 36	N° 2 TEX2 54	
1 amour»	N° 3 TEX3 10	N° 4 TEX4 11	
1 amouracher	TOUS LES TEXTES		
26 amoureuse	amour fréquence totale: 111		
8 amourusement	CLICHER SUR UN TEXTE (ou sur TOUS) pour y repérer les contextes du mot "amour"		
2 amoureuses	Cliquer AILLEURS dans cette fenêtre pour l'effacer		
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			

0 and_2
3 and_3
8 andouille_2
1 andré 2

Formes	Lemmes	Chercher un code	Codes	Cherch
<p>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</p> <p>Cliquer sur un mot pour l'activer</p> <p>← →</p> <p>Index hiérarchique</p> <p>Chercher un mot</p> <p>Graphique</p> <p>Éditer</p> <p>Quitter EXIT</p> <p>Retour</p> <p>Sommaire</p>	<p>111 amour</p> <p>1 amour»</p> <p>1 amouracher</p> <p>26 amoureuse</p> <p>8 amourusement</p> <p>2 amoureuses</p> <p>55 amoureux</p> <p>1 amours</p> <p>1 amphigourique</p> <p>2 amphithéâtre</p> <p>1 ample</p> <p>5 ampleur</p> <p>1 amplifia</p> <p>13 amusa</p> <p>5 amusaient</p> <p>25 amusait</p> <p>16 amusant</p> <p>2 amusante</p> <p>1 amusantes</p>	<p>N° 1 TEX1 36</p> <p>N° 3 TEX3 10</p> <p>TOUS LES TEXTES</p>	<p>N° 2 TEX2 54</p> <p>N° 4 TEX4 11</p>	<p>amour fréquence totale: 111</p> <p>CLIQUEZ SUR UN TEXTE (ou sur TOUS) pour y repérer les contextes du mot "amour"</p> <p>Cliquer AILLEURS dans cette fenêtre pour l'effacer</p>
				<p>3 and_3</p> <p>8 andouille_2</p> <p>1 andré 2</p>

Les occurrences du verbe *aimer* seront ignorées !?



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 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 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 DIOMS0/un

voyage à Versailles qu' il rencontre **Aimée**

NCMS000/voyage SP/à NPFP000/versailles PR00000/que PP3MSN0/il

d' **Arranet** , dont il tombe éperdument

Fc/, PR00000/dont PP3MSN0/il RG/éperdument

amoureux dès le premier regard . </S>

AQ0MN00/amoureux Fp/.

2 [doc#15462](#)

<S> Il doute aussi de son amour pour **Aimée**

PP3MSN0/il RG/aussi SP/de DP3CSS/son NCMS000/amour SP/pour

d' **Arranet** et pense de plus en plus à la

CC/et SP/à DA0FS0/le

Satin : " A l' amour , feu couvant

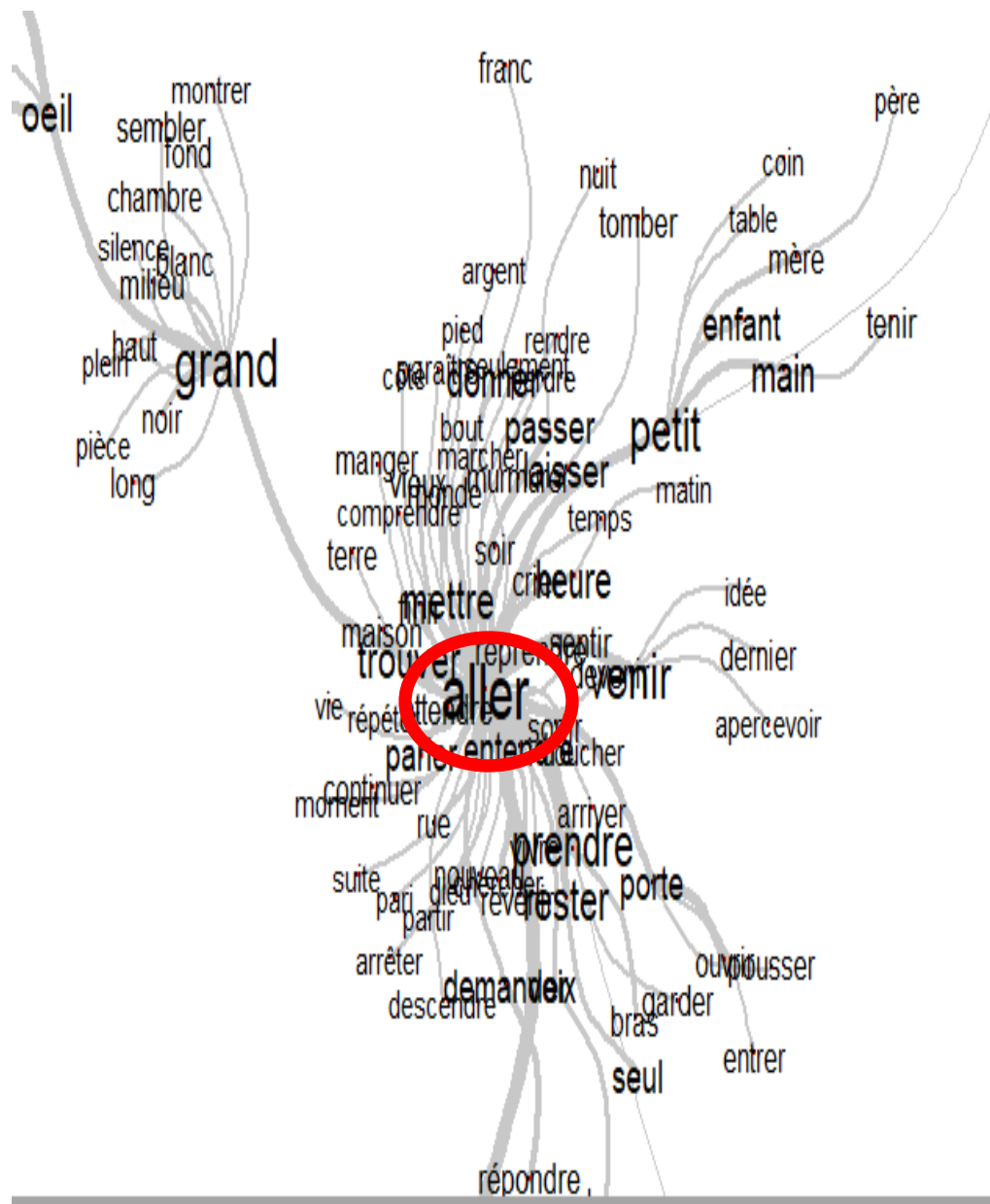
NPMS000/satin Fd/: Fe/" DA0CS0/le NCMS000/amour Fc/, NCMS000/feu VMG0000/couver

, 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





Les mots grammaticaux ne devraient pas être présentés comme des concepts caractéristiques d'un texte

aller a de nombreux sens

1. Joe va manger [Joe **is going to** eat]
2. Joe va à la maison [Joe **is going** home]
3. Comment allez-vous ? [How **are** you?]
4. Ces couleurs vont bien ensemble [these colors **match**]
5. Ce bois va de la route jusqu'à la rivière
6. Leur engagement va croissant
7. Tous ses vœux vont à la paix
8. Vous allez vite en besogne
9. Rien ne va plus
10. Sa santé va en déclinant
11. Il en va de même pour le prix du gaz
12. Allez, nous en viendrons à bout !
13. Il y va de notre santé
- 14...

La table LG1 répertorie une centaine de verbes auxiliaires (Gross 1973).

Le verbe principal n'est pas l'auxiliaire ex. :

elle n'arrête pas de lire

il ne manquera pas de pleuvoir

il devrait s'efforcer d'arrêter de fumer

NooJ - [C:\Users\Max\Documents\NooJ\fr\Lexical Analysis\LG1.dic]

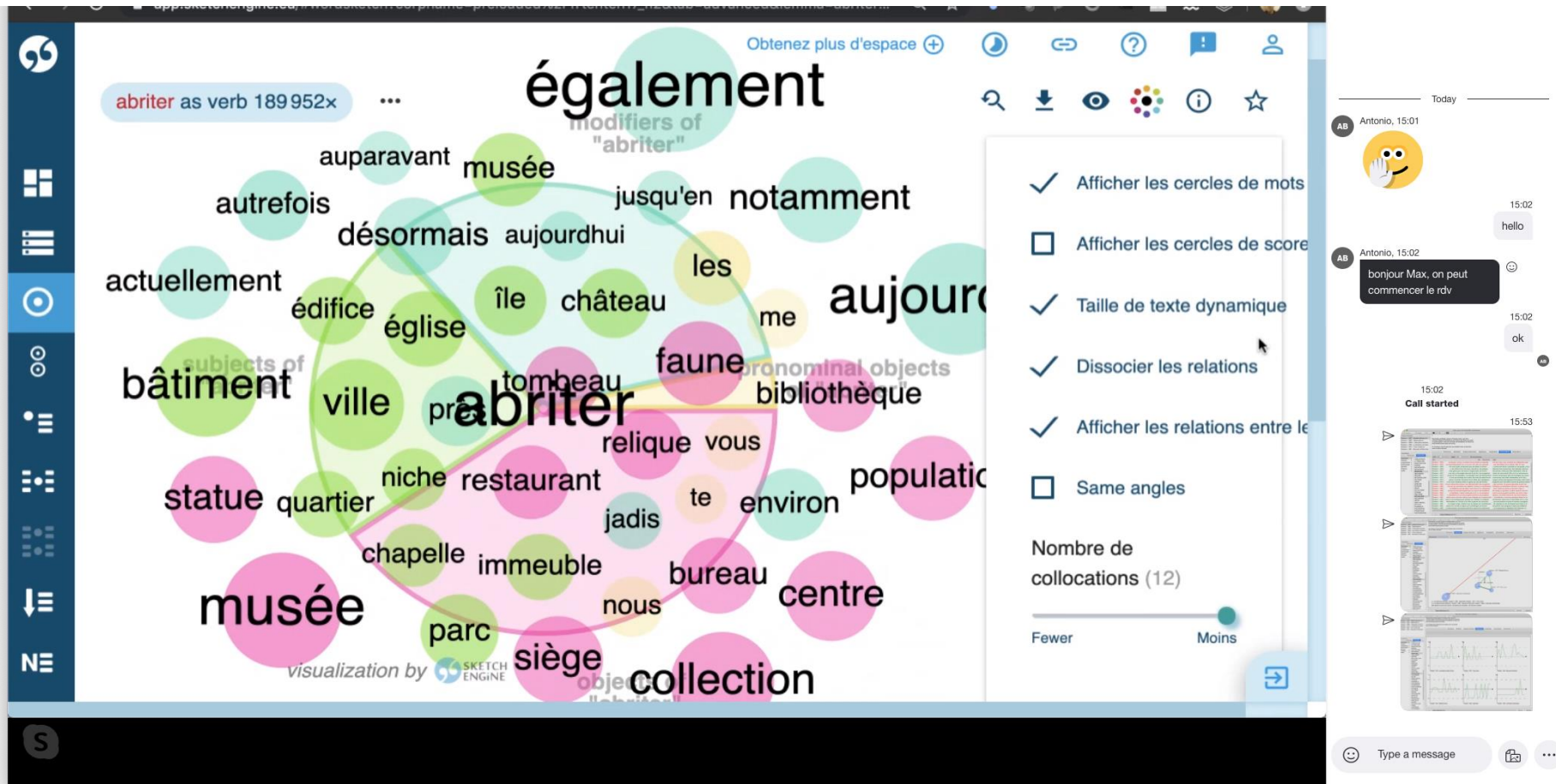
File Edit Lab Project Windows Info Max DICTIONARY

Dictionary contains 78 entries

Entrée	Catégorie	Aux	FLX	Ppv0	Prép	Spec	Tc	Tp	Vc	SynSem
achever	V	avoir	PESER		"de"	-		tc		
aller	V	être	ALLER		"jusqu'à"	-		tc		
aller	V	-	ALLER		"<E>"	-			devoir+pouvoir+savoir	
aller	V	être	ALLER	"ne"	"sans"	-		tc	devoir+pouvoir+savoir	
arrêter	V	avoir	AIMER		"de"	-		tc		N0u
arrêter	V	avoir	AIMER	"ne"	"de"	-				
arrêter	V	être	AIMER	"se"	"de"	-		tc		N0estVppW+N0u
avancer	V	être	PLACER	"se"	"à"	-		tc		
avoir	V	avoir	AVOIR		"à"	x	futur		savoir	
cesser	V	avoir	AIMER		"de"	-		tc	devoir+pouvoir+savoir	N0u
choisir	V	avoir	FINIR		"entre"	-	présent+futur			N0u
commencer	V	avoir	PLACER		"à"	-		tc	pouvoir+savoir	N0u
commencer	V	avoir	PLACER		"par"	-		tc	devoir+pouvoir+savoir	
contenter	V	être	AIMER	"se"	"de"	-	passé+présent		pouvoir+savoir	
continuer	V	avoir	AIMER		"à"	-		tc	devoir+pouvoir+savoir	N0u
continuer	V	-	AIMER		"de"	-		tc	devoir+pouvoir+savoir	
daigner	V	avoir	AIMER		"<E>"	-				
déballonner	V	être	AIMER	"se"	"de"	-		tc		N0u
décider	V	être	AIMER	"se"	"à"	-		tc	savoir	N0estVppW+N0u
défiler	V	être	AIMER	"se"	"de"	-		tc		N0u
dégonfler	V	être	AIMER	"se"	"de"	-		tc		N0u
dégrouiller	V	être	AIMER	"se"	"de"	-		tc	savoir	N0u
démerder	V	être	AIMER	"se"	"de"	-		tc	pouvoir+savoir	N0u
dépêcher	V	être	AIMER	"se"	"de"	-		tc		N0u
devoir	V	avoir	DEVOIR		"<E>"	-	passé+présent		pouvoir+savoir	
devoir	V	avoir	DEVOIR		"<E>"	-	passé+présent+futur		devoir+pouvoir+savoir	
devoir	V	être	DEVOIR	"se"	"de"	-	présent+futur		pouvoir+savoir	
efforcer	V	être	PLACER	"se"	"de"	-	futur		pouvoir+savoir	Ppv
empreser	V	être	AIMER	"se"	"de"	-		tc		N0estVppW+N0u+Ppv
être	V	avoir	ETRE		"pour"	x	futur			
finir	V	avoir	FINIR	"ne en"	"de"	-		tc		N0u
finir	V	avoir	FINIR		"de"	-		tc		N0u
finir	V	avoir	FINIR		"par"	-		tc	devoir+pouvoir+savoir	
gêner	V	être	AIMER	"ne se"	"pour"	-	présent+futur			N0estVppW
grouiller	V	être	AIMER	"se"	"de"	-		tc		N0u
hâter	V	être	AIMER	"se"	"de"	-		tc	pouvoir+savoir	N0u
hésiter	V	avoir	AIMER		"à"	-		tc		N0u
hésiter	V	avoir	AIMER	"ne"	"à"	-		tc		N0u
interrompre	V	être	ROMPRE	"se"	"de"	-		tc		N0u
lancer	V	être	PLACER	"se"	"à"	-		tc		N0estVppW+N0u
magner	V	être	AIMER	"se"	"de"	-		tc		N0u

Cancel

Le verbe *abriter* est « caractéristique » du texte... Oui mais dans quel sens ?



Dictionnaire *Les Verbes Français*, (Dubois, Dubois-Charlier 1997) :

- abriter1 (T11b8, P10b8) : (On)_s abrite (Paul)_O (de la pluie)_P (avec un parapluie) **POSITIF**
- abriter2 (T1101, P1001) : (On)_s abrite (des réfugiés)_O (chez soi) **POSITIF**
- abriter3 (T3100) : (Cet immeuble)_s abrite (les services du ministère)
- abriter4 (P10b1) : (On)_s s'abrite (des ennuis)_P (derrière son chef) **NEGATIF**
- abriter5 (T13b8) : (On)_s abrite (le port)_O (des vagues)_P (avec des digues)

Principe des méthodes empiriques

- Les outils d'analyse de texte qui suivent une méthode empirique « trouvent » leurs résultats en se référant à des corpus de référence, dits « d'apprentissage », en fait des antisèches :



- Cela fonctionner (à peu près) bien si :
 1. l'antisèche est fiable : non, malheureusement
 2. l'antisèche contient la même question : impossible
 3. le système sait quoi recopier : les formes graphiques ne sont pas pertinentes : les textes sont constitués d'unités linguistiques

1. Exemple : le corpus *Penn Treebank* n'est pas fiable

Battle-tested/NomPropreSingulier Japanese/NomPropreSingulier industrial/Adjectif managers/NomPluriel here/Adverbe always/Adverbe **buck/Verbe up/Préposition** nervous/Adjectif newcomers/NomPluriel with/Préposition the/Déterminant tale/NomSingulier of/Préposition **the/Déterminant first/Adjectif of/Préposition their/PronomPossessif** countrymen/NomPluriel to/to visit/Verbe Mexico/NomPropreSingulier,/, **a/Déterminant boatload/NomSingulier of/Préposition samurai/NomPluriel warriors/NomPluriel blown/ParticipePassé ashore/Adverbe 375/Nombre years/NomPluriel ago/Adverbe ./.** **From/Préposition the/Déterminant beginning/NomSingulier ,/, it/PronomPersonnel took/VerbePassé** a/Déterminant man/NomSingulier with/Préposition extraordinary/Adjectif qualities/NomPluriel **to/TO** succeed/Verbe in/Préposition Mexico/NomPropreSingulier ,/, ”/” says/VerbePrésent3èmeSingulier **Kimihide/NomPropreSingulier Takimura/NomPropreSingulier ,/, president/NomSingulier of/Préposition Mitsui/NomPluriel group/NomSingulier 's/Possessif Kensetsu/NomPropreSingulier Engineering/NomPropreSingulier Inc./NomPropreSingulier unit/NomSingulier ./.**

buck up = encourager
mais buck = ruer

Les corpus d'apprentissage plus récents, ex. COCA, OANC, sont aussi bourrés de fautes, cf. (Silberztein 2016)

2. Impossible d'avoir un corpus de référence suffisamment grand pour couvrir tous les cas d'utilisation

- Les projets actuels : 1 Billion Word Language Model Benchmark
- 1 milliard de mots = seulement 10^8 phrases
- Une phrase simple (sans complément circonstanciel) typique :

Our colleagues have offered a beautiful bouquet to the head of the institute.

Nos collègues ont offert un beau bouquet de fleurs au directeur de l'institut.

$$10^3 \times 10^4 \times 10^2 \times 10^2 \times 10^3 \times 10^4 \times 10^4 \times 10^2 \times 10^3 \times 10^4 \times 10^2 \times 10^3 \times 10^4 \\ = 10^{38} \text{ phrases potentielles construites avec cette structure}$$

- Le nombre de phrases est en fait en nombre infini
- Le nombre de structures potentielles pour des phrases de 13 mots :
<DET> <N+Hum> <V> <V+Dat> <DET> <A> <N+Conc> <PREP> <DET> <N+Hum> <PREP> <DET> <N>
(10^{13} combinaisons possibles ; aucun corpus ne contient 10^{13} phrases de 13 mots)

3. Traiter des formes graphiques plutôt que les unités linguistiques

- Les outils recherchent des séquences de formes graphiques
- Par exemple, pour désambiguïser la forme graphique *place* (verbe ou nom), on recherche dans le corpus d'apprentissage ses contextes, ex. :
 - ... Déterminant *take* ... => *take* est un nom
 - ... *take* Déterminant... => *take* est un verbe
- Mais on peut trouver n'importe quel type de mot avant et après *take*, que ce soit un nom ou un verbe
- Si le corpus ne peut pas être suffisamment grand, il ne contiendra pas tous les cas de contextes de *take*, donc l'outil produira des fautes systématiques.
- Si le corpus contient tous les cas (pages suivantes), il produira des résultats "statistiquement" optimisés, i.e., pas fiables.

Utiliser le contexte local d'une forme pour la désambiguïser

la forme *place* : verbe ou nom ? Avant cette forme, on peut trouver :

Adjective + Noun	It occupies a huge place in the lives of people.
Adjective + Verb	Players feeling lucky place their bet...
Adverb + Noun	I started to notice how quickly places that I found interesting...
Adverb + Verb	They quickly place the cross on the wall.
Determiner + Noun	... cover the bet and the place and win are how you make ...
Determiner + Verb	The place bet payoffs are lower than a win bet...
Conjunction + Noun	The time and place that gave me life...
Conjunction + Verb	How to choose and place art.
Noun + Noun	Her birth place was Basel.
Noun + Verb	Authors place their books under the spotlight.
Preposition + Noun	The Places API lets you search for place information.
Preposition + Verb	A purchase manager has decided to place an order.
Pronoun + Noun	When you give them place they will take yours.
Pronoun + Verb	They place him in a tomb of stone.

Utiliser le contexte local d'une forme pour la désambiguïser

la forme *place* : verbe ou nom ? Après cette forme, on peut trouver :

Verb + Adjective	Swedish women place nice significance on the value of schooling.
Noun + Adjective	It is a place nice to visit.
Verb + Adverb	They place very high demand on their suppliers.
Noun + Adverb	... that have snapped into place suddenly over the past month.
Verb + Determiner	you can use the system to place that order.
Noun + Determiner	I stumbled across this gem of a place the weekend before.
Verb + Conjunction	make money no matter what they place but only first place is.
Noun + Conjunction	LED walls have their place but there's still room for projectors.
Verb + Noun	It is allowed to place chickens in a barn with cattle.
Noun + Noun	These wedding place card ideas are sure to get ideas flowing.
Verb + Preposition	... debris which local drivers often place behind their cars.
Noun + Preposition	Everyone must visit this place in their life.
Verb + Pronoun	People in our country place those with specialized skills.
Noun + Pronoun	we could sit on another empty table near the place they showed us.

Conclusion

- Des résultats spectaculaires et magiques, mais pas fiables et non contrôlables, donc qu'on ne peut pas améliorer
- NooJ suit l'approche linguistique

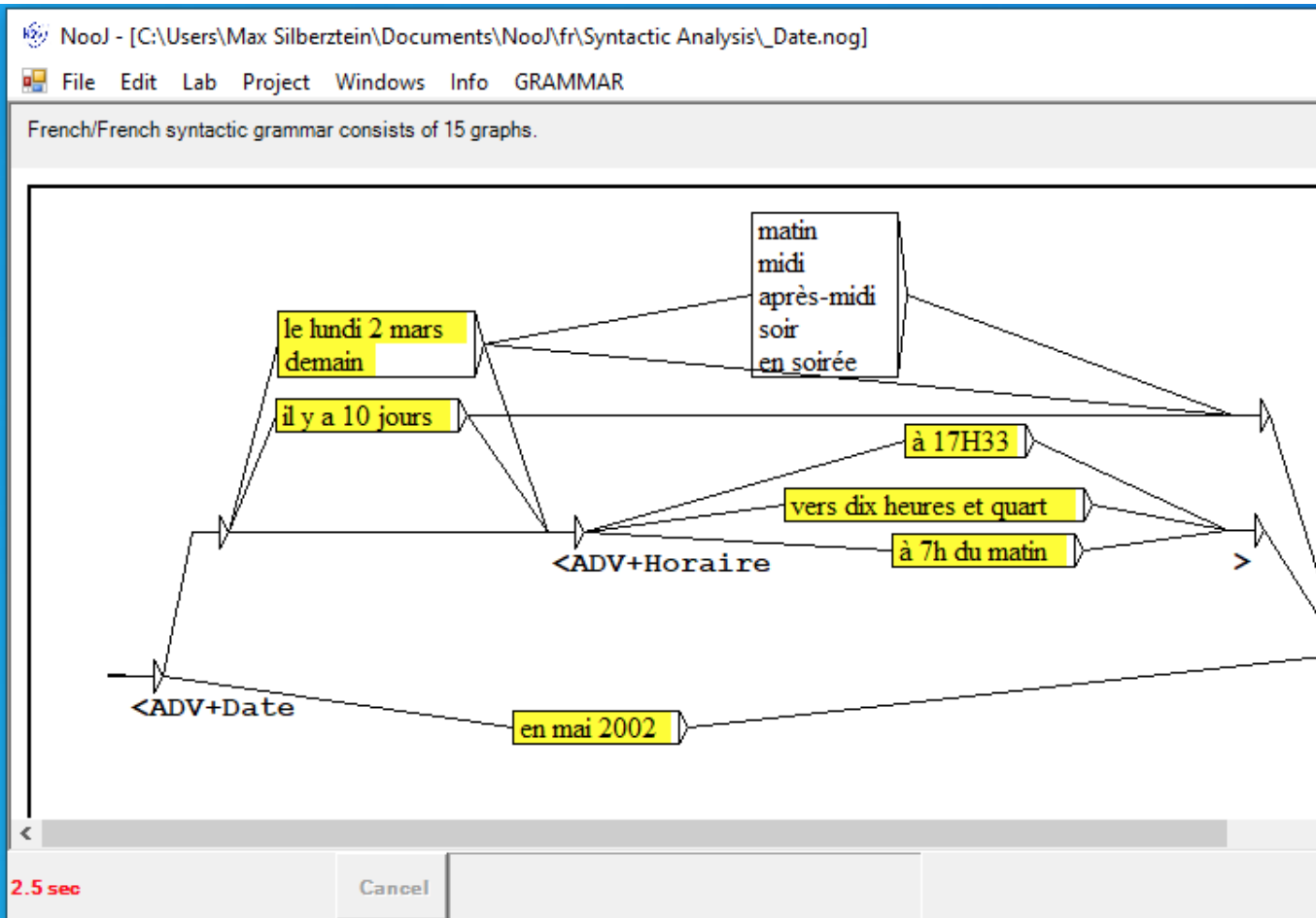
NooJ

- Historique :
 - INTEX (1993-2002) : <https://intex.univ-fcomte.fr>
 - Incident Unitex : <https://nooj.univ-fcomte.fr/intex-and-unitex.html>
 - NooJ (2002-) : <https://nooj.univ-fcomte.fr>
- Les outils pour analyser un texte :
 - Application WEB et Web API : <https://webnooj.univ-fcomte.fr>
 - Application Windows et Java (macos, linux) : <https://nooj.univ-fcomte.fr>
- Construire des concordances intelligentes
 - requêtes simples, ex. rechercher quelques mots
 - requêtes de plus en plus complexes, ex. rechercher des thèmes
 - grammaires, ex. utiliser le contexte pour lever des ambiguïtés
- Analyses statistiques : analyse normale, score standard, analyse factorielle

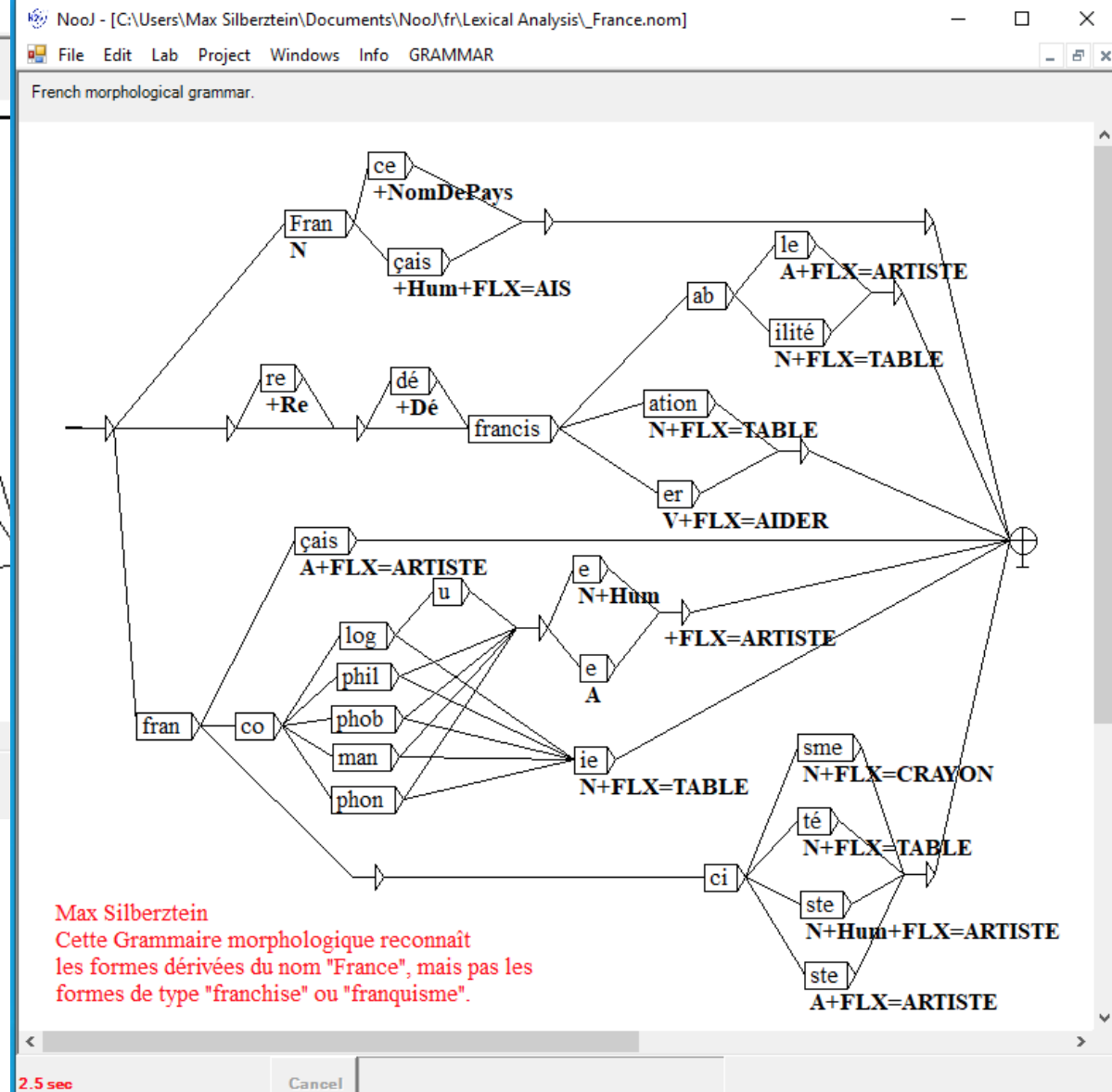
NooJ

- Comment des thèmes tels que « alcoolisme », « amour », « mort », etc. sont utilisés dans un corpus de romans ? Est-ce que le thème de la « mort » est-il plus présent dans un certain chapitre, ou alors à la fin de la vie d'un auteur ? Comment les femmes « exotiques » sont-elles décrites dans la littérature anglaise du XVIIIe ?
- Analyser/annoter les thèmes « immigration », « éducation » dans les discours d'une personnalité politique : est-elle vraiment alignée avec son parti ?
- Les phrases conjuguées au futur dans un roman, ou à la première personne chez un auteur, l'utilisation d'adverbes (inutiles) dans des articles scientifiques
- Les dénégations (« Ce n'est pas... »), la communion (« elle et moi »), l'agency (« je veux partir ») dans des interviews non directives
- Annoter les termes négatifs (« mauvais », « inutile ») dans des discours officiels ou dans des forums à propos d'un nouveau produit

Décrire des termes, expressions et structures syntaxiques



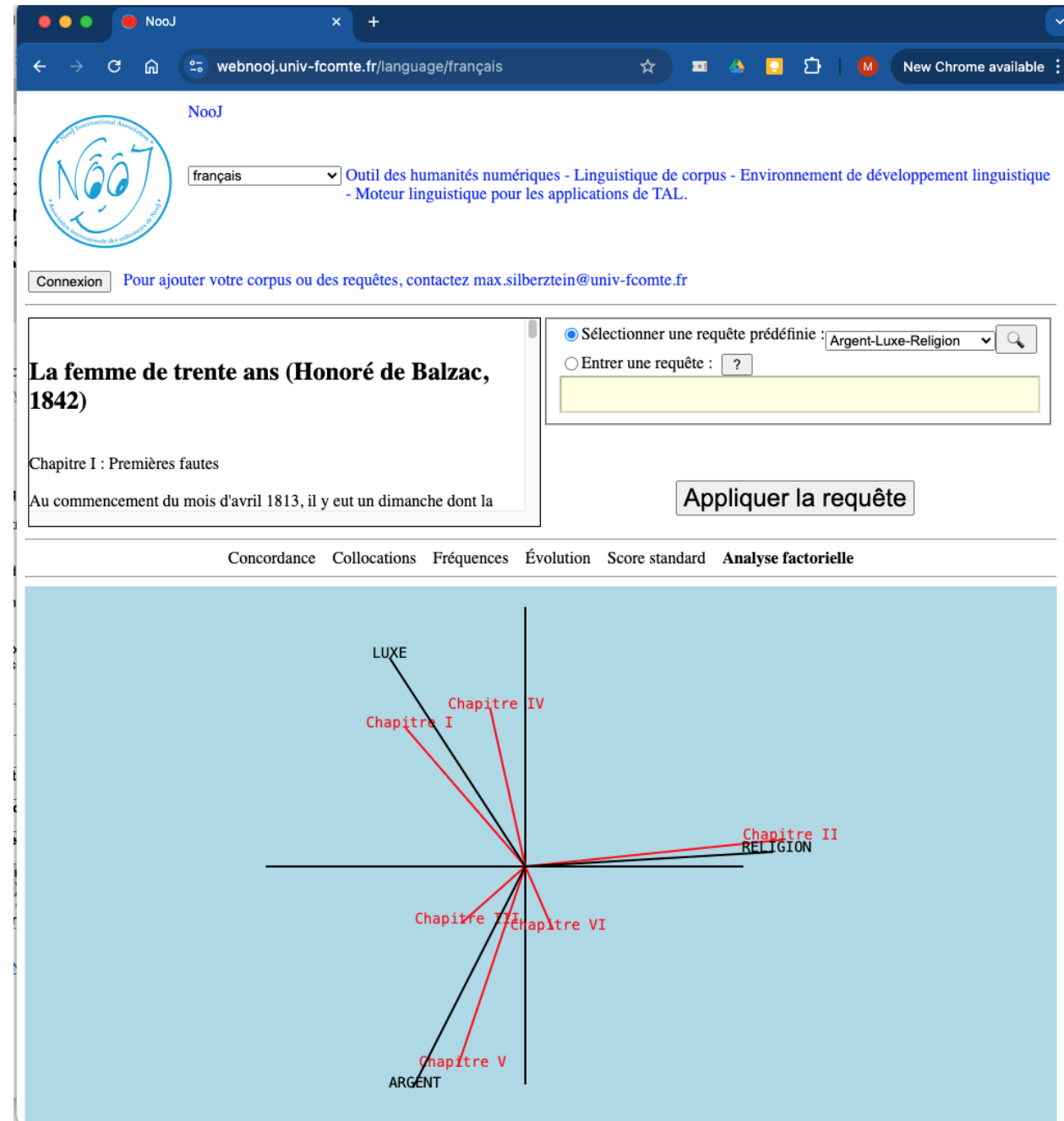
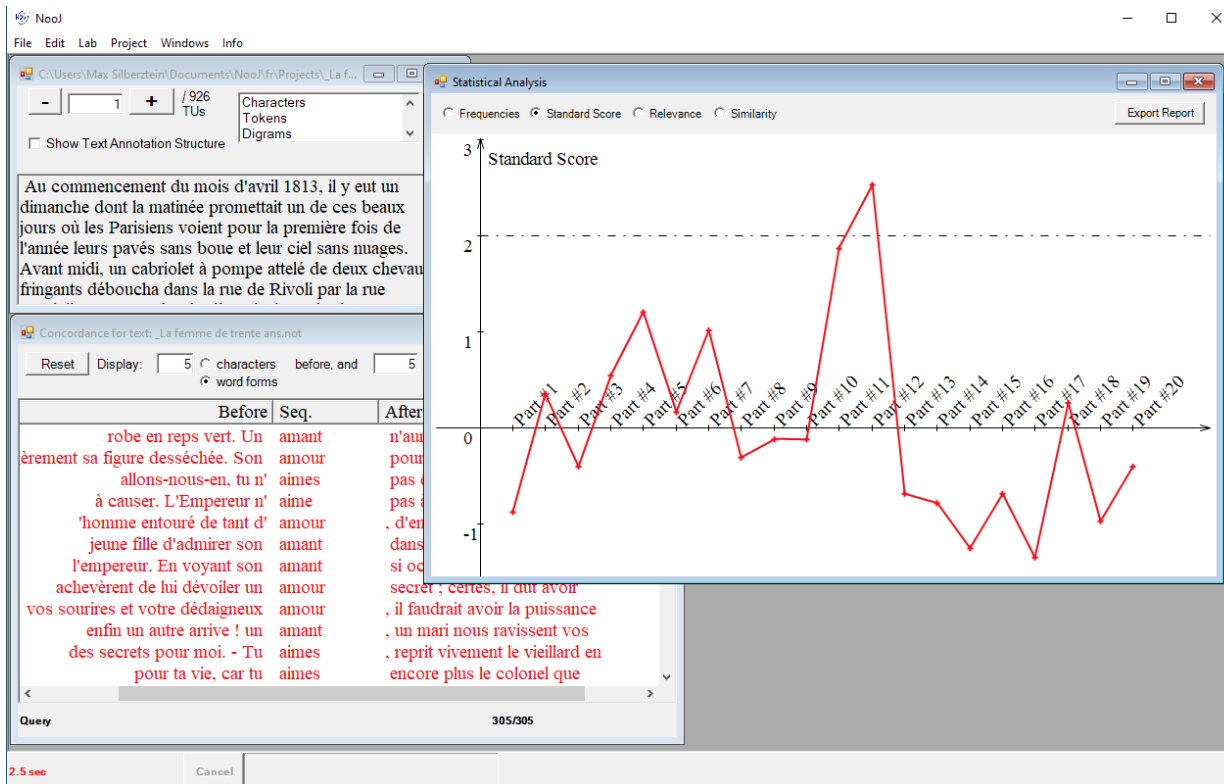
Décrire des familles de termes, entités nommées, champs lexicaux



Application WEB

Analyses statistiques appliquées à des éléments linguistiques plutôt qu'à de simples formes graphiques

Application Windows



Références

- Silberztein, M. 2015. La formalisation des langues : l'approche de NooJ. ISTE: Londres (425 pages)
- Silberztein, M. 2003-. NooJ Manual. Téléchargeable à partir de <https://nooj.univ-fcomte.fr>
- Plus de 500 références mentionnées à la page <https://nooj.univ-fcomte.fr/references.html>
- Tutoriels sur YouTube en arabe, biélorusse, croate, anglais, indonésien, italien et espagnol