



Approche technico-fonctionnelle avec NOOJ : Le cas du traitement spatio-temporel de la presse vietnamienne

*ISIC Nicolas BOFFO
Membre du bureau de la CIMI*



@communaute_cimi

Intervention du 15 janvier 2021
14h00-14H40

SEMAINE NOOJ 2021 INALCO



I. Introduction

II. Axes théoriques

III. Approche technico-fonctionnelle

IV. Le vietnamien et la temporalité

V. Problématique – Hypothèses

VI. Méthodologie

VII. Traitement spatio-temporel dans une approche technico-fonctionnelle : le cas de la presse vietnamienne.

VIII. Exemple de DATAVIZ2

IX. Conclusion





I) INTRODUCTION

La CIMI a choisi d'expérimenter NOOJ dans une approche technico-fonctionnelle sur une langue peu dotée (vietnamien) avec l'appui de l'ingénieur de Recherche Philippe LAMBERT de l'Université de Lorraine.

NOOJ est-il accessible pour des utilisateurs travaillant sur des données langagière avec un besoin de DATAVIZ et Cartographie ?

Quelques mots-clés de notre expérimentation :

Formalisation

Approche technico-fonctionnelle

Déploiement rapide

Cadres spatio-temporels

TAL

Visualisation spatio-temporelles

Langue peu dotée : vietnamien





II) AXES THEORIQUES LINGUISTIQUES

Linguistiques :

Une approche praxématique.

Une approche sociolinguistique (William LABOV).

Les cadres du discours et leurs frontières » M. CHAROLLES





II) AXES THEORIQUES LINGUISTIQUES

Formalisation de la temporalité :

Relations d'Allen.

Etudes logiques du temps et de l'aspect J.L DESCLES.

Approche de J. BRES (cf. « Temps de l'indicatif en Français »).

Sémantique de la Temporalité de GOSSELIN.

Travaux de Delphine BATISTELLI.

S-Langages de Sylviane D. SCHWER.

TimeML.

Description de Danh Thanh Do-Hurinville dans « Temps, Aspect et Modalité en vietnamien » (2009).

Description de Kim Phuong dans « Ngu phap tieng viet nhung van de thoi, the »(2008).

Approche linguistique de Cao Xuan Hao.





II) AXES THEORIQUES LINGUISTIQUES

TAL :

- NOOJ de M. SILBERZTEIN.
- Travaux de Delphine BASTISTELLI.





III) APPROCHE TECHNICO-FONCTIONNELLE

Définition : Qui prend en compte la dimension technique : types de données à traiter, support de la données, architecture à mettre en place, etc.

ainsi que la dimension fonctionnelle :

Missions à réaliser => veille internet, analyse documentaire, traduction automatique, etc.

ZONE GRISE : pas de définition précise dans la bibliographie.

On parle beaucoup de consultant technico-fonctionnel mais pas d'approche.





III) APPROCHE TECHNICO-FONCTIONNELLE

Définition : Qui prend en compte la dimension technique : types de données à traiter, support de la données, architecture à mettre en place, médias, etc.

ainsi que la dimension fonctionnelle :

Missions à réaliser => archive, veille internet, analyse documentaire, traduction automatique, etc.





IV) LE VIETNAMIEN ET LA TEMPORALITE

- Le vietnamien est un langue à tons et monosyllabiques.
- En vietnamien il n'y a pas de temps verbaux.
- Il y a du temps, de l'aspect et de la modalité.
- Il y a des marqueurs aspectuo-temporels et modaux.
- Il y a aussi des Cadres Temporels (selon la déf. par M. Charolles)





IV) LE VIETNAMEIEN ET LA TEMPORALITE

Illustration :

Đã, đang et sẽ ne sont pas des marqueurs temporels mais aspectuels (Cao Xuan Hao (1998), Danh Thanh Do-Hurinville (2008)).

Ex:

Thẻ đã mua, xin miễn đổi lại. (Once bought, this card is not exchangeable.)

đã = action is accomplished/finished.

Hôm qua khi tôi đang xem TV thì điện thoại reo. (Yesterday, when I was watching TV, the phone rang.)

đang = action is “during”.

Nêu bạn vượt đèn đỏ thì bạn sẽ bị phạt. (Si vous brûlez le feu rouge, vous êtes pénalisé.)

sẽ = action is after.





IV) LE VIETNAMIEN ET LA TEMPORALITE

Les cadres temporels :

Năm ngoái (IF)

Từ 12/09 đến 24/10 (IF)

Một ngày (IF)

Từ 19/09 đến 09/11 (IF)





IV) LE VIETNAMIEN ET LA TEMPORALITE

- Catégories Temporelles:

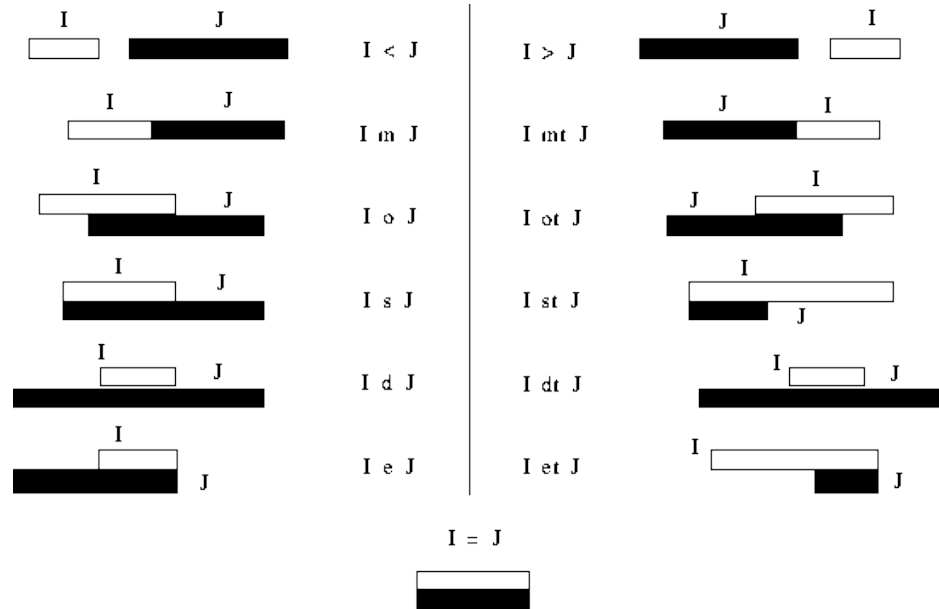
- Passé
- Present
- Futur
- Neutre

- Relations temporelles :

- Relations d'Allen

- Catégories Aspectuelles:

- Accompli (đã)
- Progression (đang)
- Iterative (V+lại)
- Terminative (xong, hết, thôi, etc.)
- Imminente (sắp)





V) PROBLEMATIQUES - HYPOTHESES

Problématique (s) :

- Comment mettre en œuvre rapidement un « système » opérationnel/fonctionnel pour un utilisateur souhaitant traiter des données langagières en vietnamien ?
- NOOJ répond-il à ce besoin technico-fonctionnel ?

Hypothèses :

- NOOJ permet de formaliser des langues peu dotées tel que le vietnamien.
- L'utilisation de grammaires locales permettra de se concentrer sur l'essentiel au niveau fonctionnel.
- Avec NOOJ nous pouvons faire de l'analyse syntaxique de surface.
- L'analyse lexico-syntaxique est privilégiée car le VN est une langue monosyllabique et non-flexionnelle => possibilité de formaliser cela par des Automates.
- Les extractions (sorties) de NOOJ peuvent être mises en entrées pour faire de la DATAVIZ et traitement SIG.



V) PROBLEMATIQUES - HYPOTHESES

Hypothèses :

Les cadres temporels définissent les univers temporels (M. CHAROLLES dans les « Cadres du Discours »)

=> Dans *Logique du sens* (1983), R.Martin définit les univers de discours comme "l'ensemble des circonstances, souvent spécifiées sous forme d'adverbes de phrase, dans lesquelles la proposition peut être dite vraie" (p.37). Nous allons partir de cette définition qui attire immédiatement l'attention sur:

- le fait que **les univers de discours ont rapport aux circonstances des événements, procès et états dénotés dans le cours d'un texte,**
- le fait qu'**ils sont généralement introduits** et plus ou moins nettement spécifiés (catégorisés) **par des groupes syntaxiques périphériques adjoints à la phrase,**
- et le fait qu'**ils précisent les conditions** dans lesquelles la proposition qu'ils indexent peut être dite vraie ou fausse.





VI) METHODOLOGIE

Construction des grammaires temporelle du vietnamien

Travail linguistique :

- Observation et analyse de l'aspect et du temps.
- Grammaires et études linguistiques vietnamiennes de Tran Kim Phuong, D.T. Do-Hurinville, Thomson, Cao Xuan Hao, Nguyen Van Hue, etc.
- Méthodes de langue vietnamienne étrangère (Université de Hanoi et de Ho Chi Minh Ville, etc.)
- Dictionnaires de Le Kha Ke, Nguyen Trong Bau.
- Lexiques du « Trung tam bien soan dich thuat sach sai gon ».
- Coocurence : Utilisation du TRAMEUR de Paris 3.

↙
Dictionnaires MT VN>FR
Tableaux N-2 N-1 N0 N+1 N+3
Automates (NOOJ)





VI) METHODOLOGIE

Un exemple d'analyse temporelle avec la Classe Temporelle Universelle basée sur TimeML et nos analyses linguistiques

Exemple:

| Hôm | | qua | | tôi | đang | xem | TV | thì | điện thoại | reo.

Jour | | passé/révolu | | je | marqueur aspectuel « continu » | regarder | TV | alors |
téléphone | sonner.

Annotation:

<OPEN> | Hôm<N0> | | qua <N+1><ACCOMPLISHED> | <+PAST> |
| tôi | đang<-INCIDENCE><PROGRESSIVE> | xem<V> | TV | thì<MID> | điện thoại |
reo<+INCIDENCE><ACCOMPLISHED> | <CLOSE>.

Représentation graphique et SLS:



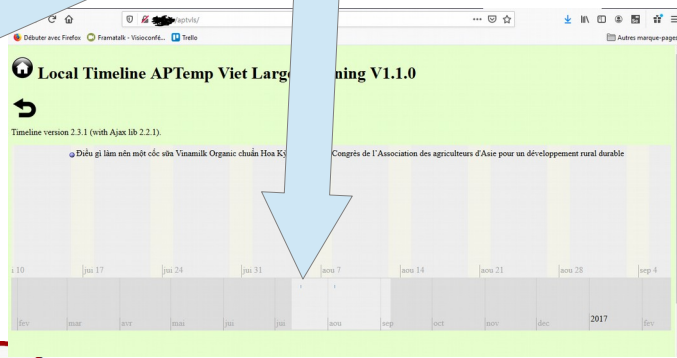
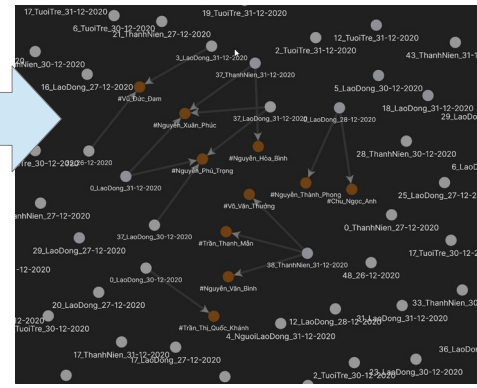
Pour cette démonstration, nous nous sommes appuyé sur analyse temporelle qui avait fait l'objet de travaux antérieurs :

- « A new model of Time Expressions Detection and Annotation in Vietnamese : The hôm case. », S. SCHWER, P. LAMBERT et N. BOFFO, IALP 2012
- « Towards a unified model of identification and description for temporal reference in Nooj : the example of day-expressions in Vietnamese. » S. SCHWER, P. LAMBERT et N. BOFFO, Nooj 2013 International Conference, June 3-5, 2013, Saarbrücken (Deutschland)
- Semaine NOOJ 2013 : Formalisation de la temporalité en vietnamien, N. BOFFO

VI) METHODOLOGIE

Constitution du corpus web
Scripts Python

NOOJ
Traitement du corpus
Grammaires
Dictionnaires

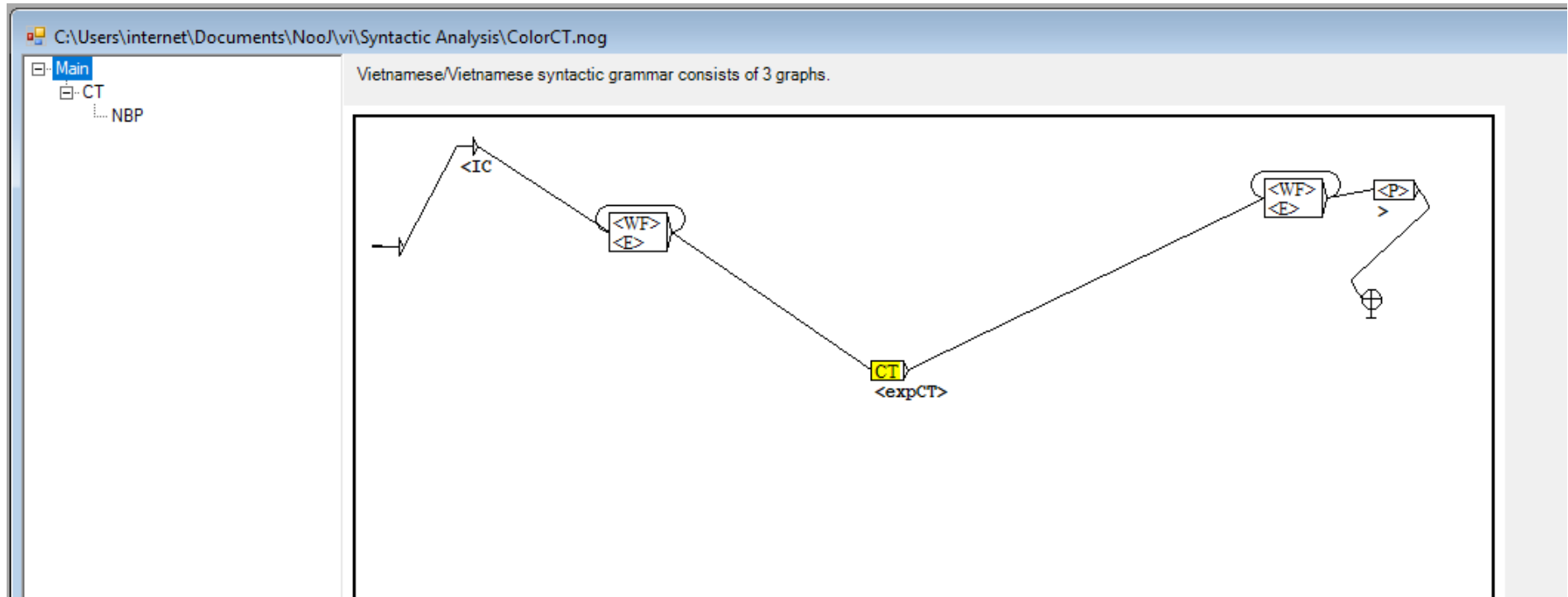




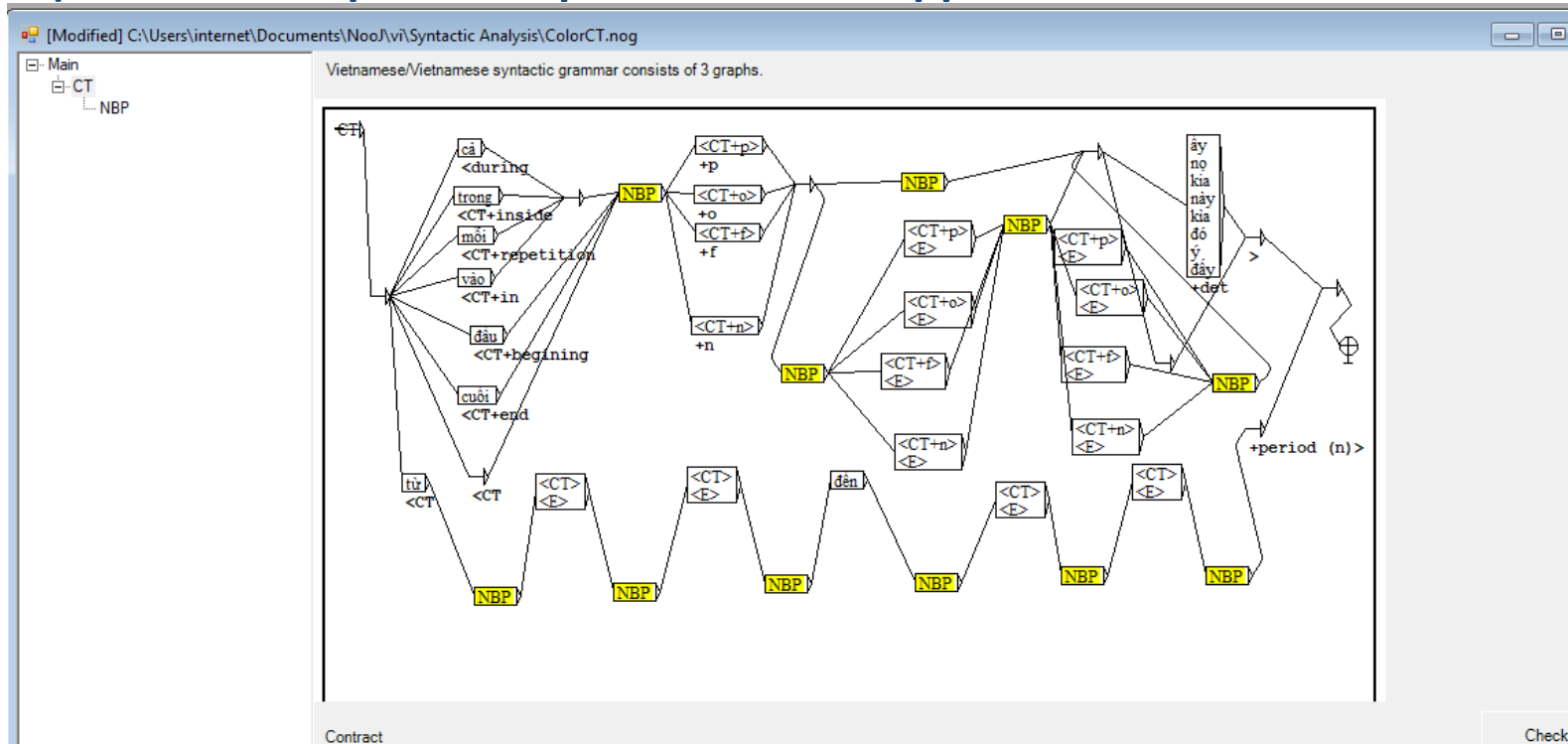
VII) Traitement spatio-temporel dans une approche technico-fonctionnelle

Exploration de la grammaire NOOJ d'extraction des cadres temporels et des informations temporelles ainsi que de ses dictionnaires.

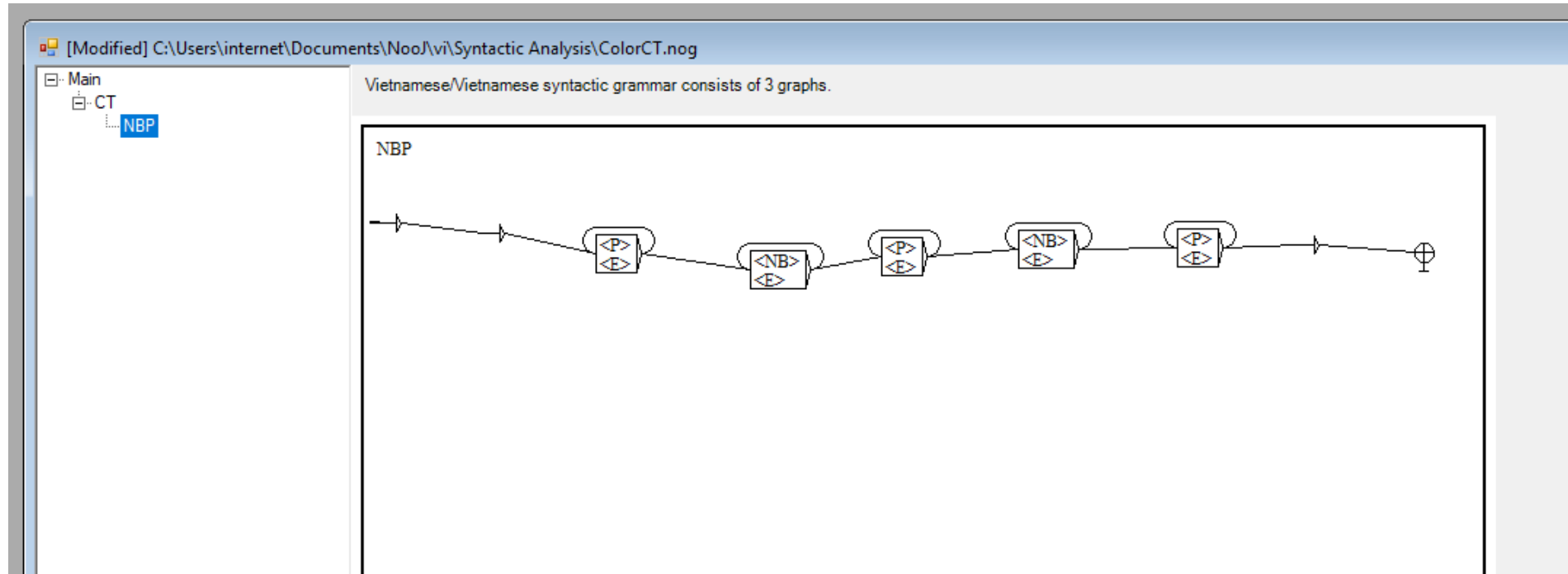
VII) Traitement spatio-temporel dans une approche technico-fonctionnelle

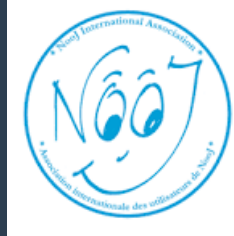


VII) Traitement spatio-temporel dans une approche technico-fonctionnelle



VII) Traitement spatio-temporel dans une approche technico-fonctionnelle





VII) Traitement spatio-temporel dans une approche technico-fonctionnelle Dictionnaire marqueurs Présen (6 entrées traduites en français)

```
C:\Users\internet\Documents\NooJ\vi\Lexical Analysis\CTO.dic
Dictionary contains 6 entries

# NooJ V2
# Dictionary
#
# Input Language is: vi
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' and properties' definition files (
# Special Command: #use properties.def
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB (u
#                   +FLX= (inflectional paradigm) +DRV= (derivational paradigm)
#
# Special Characters: '\ ' "' '+' ',' '#' ' ' '
#
hôm nay,CT+o+FR=aujourd'hui
bây giờ,CT+o+FR=maintenant
lúc nay,CT+o+FR=à présent
ngày nay,CT+o+FR=de nos jours
hiện nay,CT+o+FR=de nos jours
trong hoàn cảnh hiện nay,CT+o+FR=par le temps qui court
```

VII) Traitement spatio-temporel dans une approche technico-fonctionnelle Dictionnaire marqueurs Neutre (188 entrées traduite en français)

```
C:\Users\internet\Documents\Nooj\vi\Lexical Analysis\CTN.dic
Dictionary contains 188 entries

#
# +FLX= (inflectional paradigm) +DRV= (derivational paradigm)^
#
# Special Characters: '\ ' "' '+' ',' '# ' ' '
#
trước hôm,CT+n+FR=la veille (de)
một thời đại,CT+n+FR=une ère
kỷ nguyên,CT+n+FR=une ère
mười ngày,CT+n+FR=une décade
ba tháng,CT+n+FR=un trimestre
quý,CT+n+FR=un trimestre
một thời kỳ,CT+n+FR=une époque
thời đại,CT+n+FR=une époque
một thế kỷ,CT+n+FR=un siècle
một thời gian,CT+n+FR=une durée/une période
nửa năm,CT+n+FR=un semestre
sáu tháng,CT+n+FR=un semestre
thiên kỷ,CT+n+FR=le millénaire
bình minh,CT+n+FR=l'aube
rạng đông,CT+n+FR=l'aube
hoàng hôn,CT+n+FR=le crépuscule
thời gian,CT+n+FR=le temps
khoảng thời gian,CT+n+FR=un (certain) temps
lúc,CT+n+FR=un moment/un instant
giây,CT+n+FR=la seconde
phút,CT+n+FR=la minute
giờ,CT+n+FR=l'heure
ngày,CT+n+FR=le jour/la journée
```



VII) Traitement spatio-temporel dans une approche technico-fonctionnelle

Dictionnaire marqueurs Futur (14 entrées traduite en français)

```
C:\Users\internet\Documents\Nooj\vi\Lexical Analysis\CTF.dic
Dictionary contains 14 entries

#
# Input Language is: vi
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' and properties' definition files
# Special Command: #use properties.def
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB
#                   +FLX= (inflectional paradigm) +DRV= (derivational paradigm)
#
# Special Characters: '\ ' ' ' ' + ' ' , ' ' # ' ' '
#

ngày mai,CT+f+FR=demain
ngày kia,CT+f+FR=après-demain
ngày mốt,CT+f+FR=après-demain
sáng mai,CT+f+FR=demain matin
tối mai,CT+f+FR=demain soir
đêm mai,CT+f+FR=demain soir
sáng ngày kia,CT+f+FR=après demain matin
tối ngày kia,CT+f+FR=après demain soir
sáng hôm sau,CT+f+FR=le lendemain matin
tối hôm sau,CT+f+FR=le lendemain soir
tuần tới,CT+f+FR=la semaine prochaine
tháng tới,CT+f+FR=le mois prochain
```


VII) Traitement spatio-temporel dans une approche technico-fonctionnelle Dictionnaire marqueurs Passé (14 entrées traduites en français)

```
C:\Users\internet\Documents\NooJ\vi\Lexical Analysis\CTP.dic
Dictionary contains 12 entries

# NooJ V2
# Dictionary
#
# Input Language is: vi
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' and properties' definition files (
# Special Command: #use properties.def
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB (u
#                   +FLX= (inflectional paradigm) +DRV= (derivational paradigm)
#
# Special Characters: '\ ' "' '+' ' ' ' #' ' ' '
#
hôm qua,CT+p+FR=hier
hai hôm trước,CT+p+FR=l'avant-veille (de)
đêm qua,CT+p+FR=hier soir
tối qua,CT+p+FR=hier soir
đêm trước,CT+p+FR=la nuit dernière
tối hôm trước,CT+p+FR=la nuit dernière
tối hôm trước,CT+p+FR=la veille au soir
sáng hôm trước,CT+p+FR= la veille au matin
tuần trước,CT+p+FR=la semaine passé
tháng trước,CT+p+FR=le mois passé
năm ngoái,CT+p+FR=l'an passé
năm ngoái cũng vào thời điểm này,CT+p+FR=l'an dernier à pareille époque
```



VII) Traitement spatio-temporel dans une approche technico-fonctionnelle

Test sur la biographie de Ho Chi Minh

Concordance

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

Text	Before	Seq.
		Cuộc đời của Chủ tịch Hồ Chí Minh là một cuộc đời trong sáng cao đẹp của một người cộng sản vĩ đại, Chủ tịch Hồ Chí Minh (lúc nhỏ tên là Nguyễn Sinh Cung khi đi học là Nguyễn Tất Thành, trong nhiều năm hoạt động cách mạng lấy tên là Nguyễn Ái Quốc và nhiều bí danh, sinh ngày 19 tháng 5 năm 1890 ở Kim Liên huyện Nam Đàn, mất ngày 2 tháng 9 năm 1969 tại Hà Nội. Với tình cảm yêu nước thương dân vô hạn, năm 1911 Người đã rời Tổ quốc đi sang phương Tây để tìm con đường giải phóng dân tộc. Cuối năm 1917, Người từ Anh trở lại Pháp tiếp tục hoạt động trong phong trào Việt kiều và phong trào công nhân Pháp. Cuối năm 1917, Người từ Anh trở lại Pháp tiếp tục hoạt động trong phong trào Việt kiều và phong trào công nhân Pháp. Năm 1919, lấy tên là Nguyễn Ái Quốc, Dưới ảnh hưởng của Cách mạng Tháng Mười Nga năm 1917 và Luận cương của Lenin về vấn đề dân tộc và thuộc địa, tháng 12 năm 1920, Nguyễn Ái Quốc tham dự đại hội lần thứ XVIII Đảng Xã hội Pháp và Người bỏ phiếu tán thành Đảng gia nhập Quốc tế III, Người khẳng định con đường cách mạng giải phóng dân tộc trong thời đại mới là con đường của chủ nghĩa Mác-

Query 51/51



VII) Traitement spatio-temporel dans une approche technico-fonctionnelle

Test sur la biographie le corpus santé et corpus_Phil

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

Text	Before	Seq.
0_ThanhNien_31-12-2020.not		yêu quý, gần bó và không ngừng tiếp sức cho tờ báo.
0_ThanhNien_31-12-2020.not		Cùng lúc là việc tích cực tham gia nhiều chương trình.
0_ThanhNien_31-12-2020.not		Hội hướng vào các đối tượng thanh niên. 35 năm qua,
0_ThanhNien_31-12-2020.not		nhà báo Nguyễn Quang Thông khẳng định trong năm 2021 và các năm tiếp theo.
0_ThanhNien_31-12-2020.not		nhà báo Nguyễn Quang Thông khẳng định trong năm 2021 và các năm tiếp theo.
0_TuoiTre_27-12-2020.not		Theo báo cáo của Văn phòng Hội đồng Giáo sư Nhà nước, năm 2020,
0_TuoiTre_27-12-2020.not		Giáo sư trẻ nhất năm 2020 là ông Lê Anh Vinh (
0_TuoiTre_27-12-2020.not		Các phó giáo sư trẻ năm nay là ông Phạm Chiến Thắng (
0_TuoiTre_27-12-2020.not		chất lượng ứng viên năm 2020 khá tốt
0_TuoiTre_27-12-2020.not		phó giáo sư năm 2020.
0_TuoiTre_30-12-2020.not		ngoại ở Tehran ngày 8-1 - Ảnh:
0_TuoiTre_30-12-2020.not		của Hãng hàng không quốc tế Ukraine đã bị bắn rơi ngay sau khi vừa cất cánh khỏi sân bay tại Tehran ngày 8-
0_TuoiTre_30-12-2020.not		Ba ngày sau thảm kịch.
0_TuoiTre_31-12-2020.not		Vướng mắc này đã tồn tại suốt 16 năm qua.
0_TuoiTre_31-12-2020.not	(EC) ngày 31-12 tuyên bố động thái của Washington đã '
10_26-12-2020.not	(Tối ngày 26.12, Ban Chỉ đạo Quốc gia phòng chống dịch bệnh COVID-
10_26-12-2020.not	(sinh năm 1988, ngụ huyện Mang Thít,
10_26-12-2020.not		tỉnh Vĩnh Long). Vào ngày 15.12, người này di chuyển từ Myanmar bằng xe tải về huyện Mae Sot.
10_26-12-2020.not		Đến ngày 22.12, tiếp tục ngồi trong một thùng sau xe tải (
10_26-12-2020.not		Đến ngày 23.12,
10_26-12-2020.not		giờ sáng ngày 24.12,
10_26-12-2020.not		Đến khoảng 05 giờ sáng,
10_26-12-2020.not		sáng ngày 24.12, dừng chân tại khu vực cổng chào tỉnh Vĩnh Long,
10_26-12-2020.not		giám sát và đưa cách ly tập trung tại Trung tâm Y tế huyện Mang Thít. Sáng hôm sau, Trung tâm Kiểm soát bệnh tật Vĩnh Long thực hiện lấy mẫu xét nghiệm gửi về Viện P
10_26-12-2020.not		Kết quả xét nghiệm ngày 26.12 của Viện Pasteur TP.
10_LaoDong_28-12-2020.not		phải nhận trận thua thảm bại trước Bình Dương ở ngày khai màn Cúp Thiên Long.
10_LaoDong_28-12-2020.not		Chiều ngày 28.12, lượt trận đầu tiên của giải đấu giải giao hữu Thiên Long -
10_LaoDong_28-12-2020.not		ngay trong ngày ra quân của giải Thiên Long Cup.
10_LaoDong_28-12-2020.not		ngay trong ngày ra quân của giải Thiên Long Cup.
10_LaoDong_28-12-2020.not		bàn thắng của Hồng Phước (phút 11), Tô Văn Vũ (
10_LaoDong_28-12-2020.not	(phút 35) và Mansaray (
10_LaoDong_28-12-2020.not	(phút 57)

VII) Traitement spatio-temporel dans une approche technico-fonctionnelle

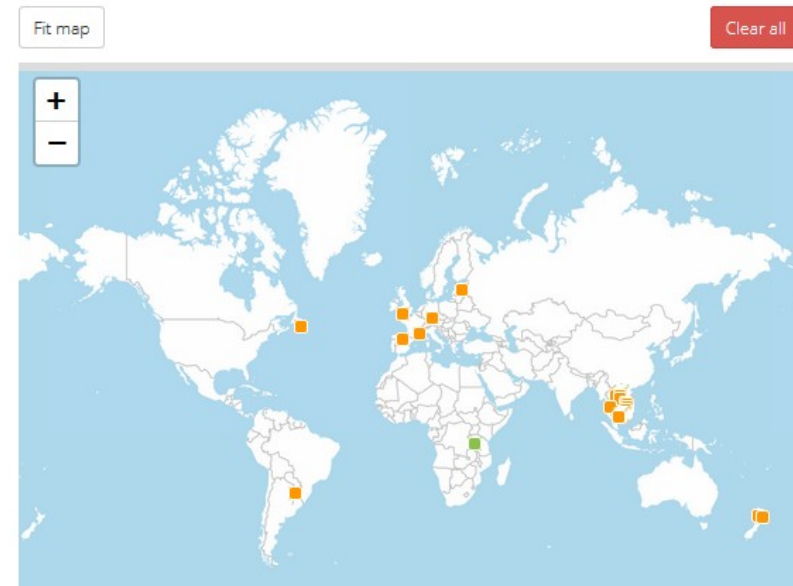
Visualisation spatiale

<https://geocode.localfocus.nl/>

3. Check the results

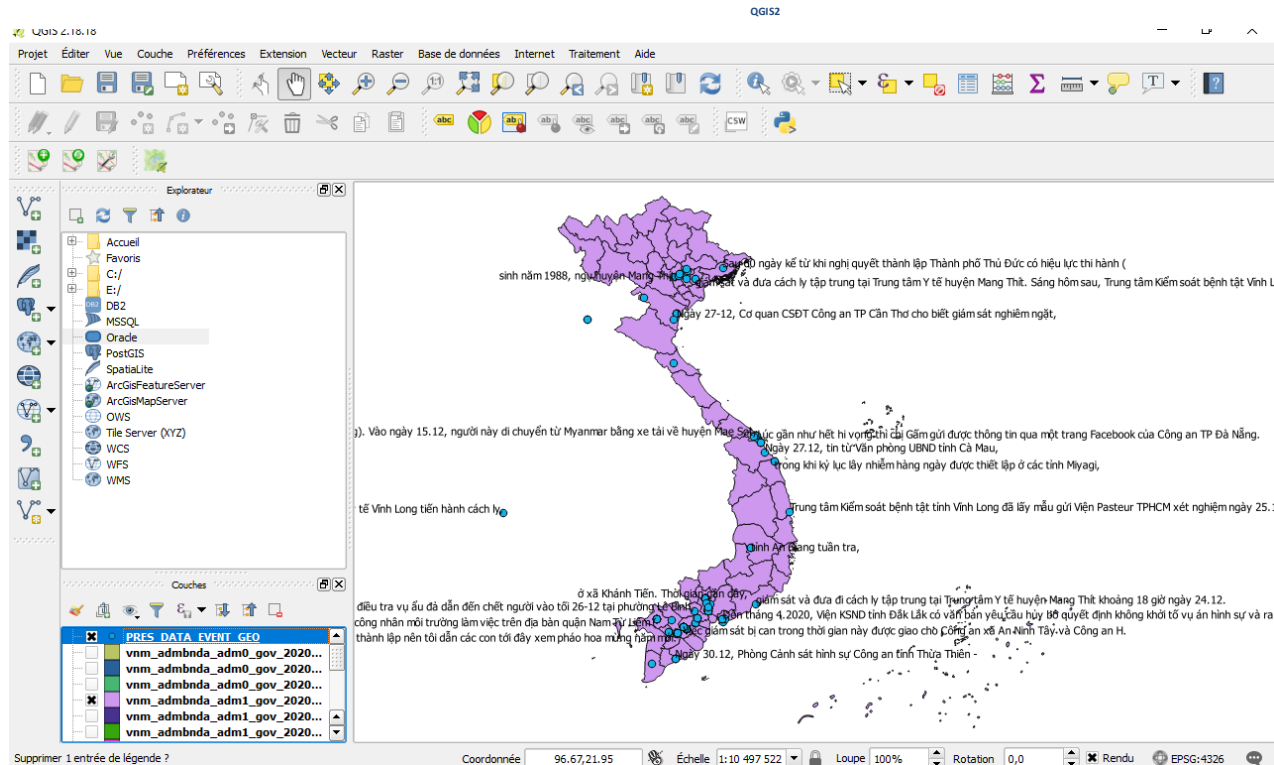
Pending 0 Success 56 Doubt 84 Failed 12

- tỉnh Vĩnh Long (Hits: 10)
- thành phố Hà Nội (Hits: 10)
- tỉnh Vĩnh Long (Hits: 10)
- huyện Mang Thít (Hits: 10)
- tỉnh Vĩnh Long (Hits: 10)
- huyện Mae Sot (Hits: 10)
- tỉnh Vĩnh Long (Hits: 10)
- huyện Mang Thít (Hits: 10)
- tỉnh Phú Thọ (Hits: 10)



VII) Traitement spatio-temporel dans une approche technico-fonctionnelle

Visualisation spatiale





VII) Traitement spatio-temporel dans une approche technico-fonctionnelle Visualisation temporelle (non contractuel, composant obsolète de 2016I)

Local Timeline APTemp-Viet Large Scanning V1.1.0

Timeline version 2.3.1 (v)

Điều gì làm nên một cốc sữa Vinamilk Organic chuẩn Hoa Kỳ?

Điều gì làm nên một cốc sữa Vinamilk Organic chuẩn Hoa Kỳ?






VII) Traitement spatio-temporel dans une approche technico-fonctionnelle Problématiques :

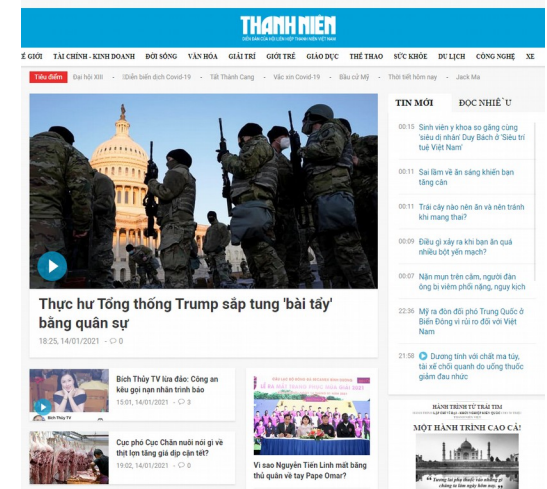
- => Encore beaucoup trop d'ambiguïtés.
- => Bruit.
- => Difficile d'évaluer le système sans corpus annoté.
- => Beaucoup de post traitements sont nécessaires (nettoyage et normalisation).



VIII) Exemple de DATAVIZ (Philippe LAMBERT)

Processus

- Extraction des articles de sites de la presse VN :
- Transformation en fichier XML 
- Traitement des nœuds XML par NooJ (<Content>)
- Extraction des termes de concordance
- Injection des termes de concordance comme #tags dans une solution de visualisation de l'information.

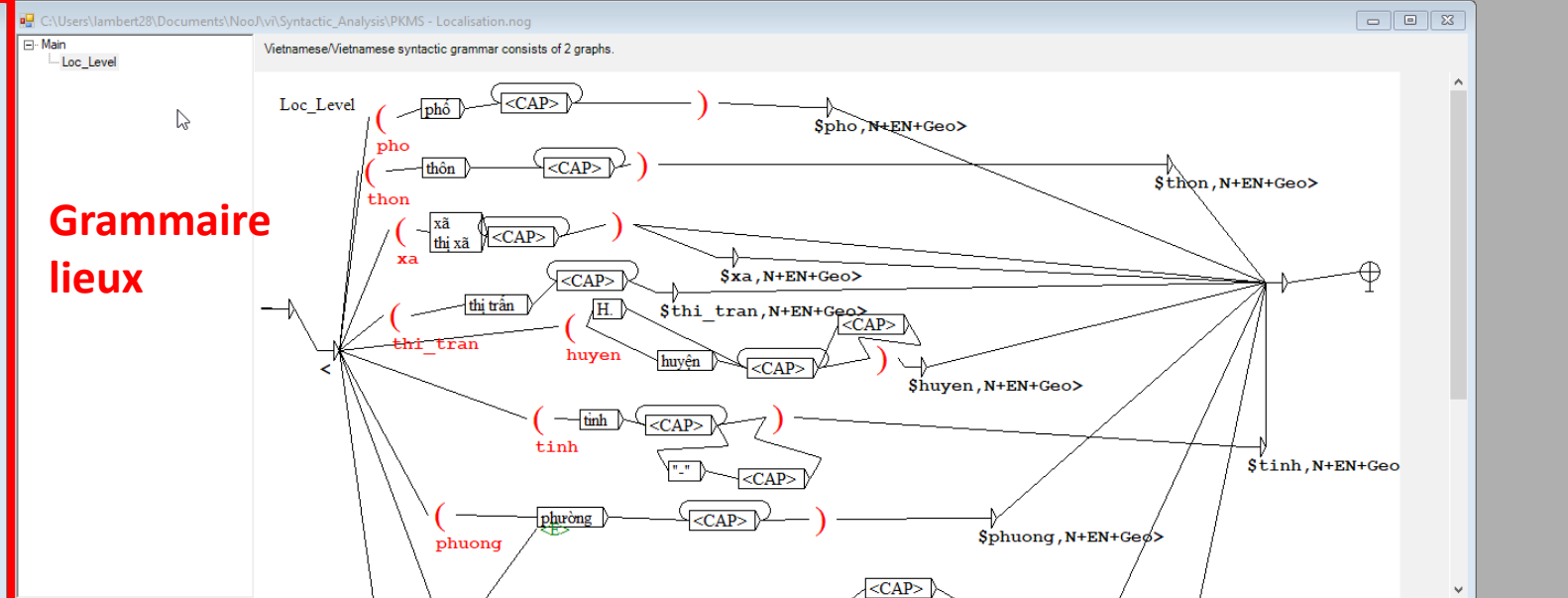


Corpus C:\Users\lambert28\Documents\NooJ\vi\Projects\corp...

Corpus Language is "Vietnamese(vi)".
Original Text File format is "Default".
Corpus consists of 450 text files
Text Delimiter is: "\n"
Corpus contains 8434 text units delimited by "\n"

File Name	Size	Last Modif.
0_26-12-2020	16180	13/01/2021
0_26-12-2020	13169	13/01/2021
0_LaoDong_27-12-2020	8779	13/01/2021
0_LaoDong_28-12-2020	20582	13/01/2021
0_LaoDong_30-12-2020	51841	13/01/2021
0_LaoDong_31-12-2020	63072	13/01/2021
0_NguoiLaoDong_30-12-2020	4063	13/01/2021
0_NguoiLaoDong_31-12-2020	1443	13/01/2021
0_ThanhNien_27-12-2020	4615	13/01/2021
0_ThanhNien_30-12-2020	5667	13/01/2021
0_ThanhNien_31-12-2020	49290	13/01/2021
0_TuoiTre_27-12-2020	12587	13/01/2021
0_TuoiTre_30-12-2020	7257	13/01/2021
0_TuoiTre_31-12-2020	9222	13/01/2021
02_26-12-2020	12051	13/01/2021
03_26-12-2020	8060	13/01/2021
035_26-12-2020	8494	13/01/2021
036_26-12-2020	17925	13/01/2021
037_26-12-2020	20565	13/01/2021
04_26-12-2020	21202	13/01/2021
05_26-12-2020	15685	13/01/2021
06_26-12-2020	8751	13/01/2021
07_26-12-2020	9391	13/01/2021
08_26-12-2020	28721	13/01/2021
1_26-12-2020	15924	13/01/2021
1_26-12-2020	18579	13/01/2021
1_LaoDong_27-12-2020	1976	13/01/2021
1_LaoDong_28-12-2020	10282	13/01/2021
1_LaoDong_30-12-2020	2020	13/01/2021
1_LaoDong_31-12-2020	21801	13/01/2021
1_NguoiLaoDong_30-12-2020	3462	13/01/2021
1_NguoiLaoDong_31-12-2020	10643	13/01/2021
1_ThanhNien_27-12-2020	5075	13/01/2021
1_ThanhNien_30-12-2020	5954	13/01/2021
1_ThanhNien_31-12-2020	6195	13/01/2021
1_TuoiTre_27-12-2020	18534	13/01/2021
1_TuoiTre_30-12-2020	16180	13/01/2021
1_TuoiTre_31-12-2020	8830	13/01/2021
10_26-12-2020	15924	13/01/2021
10_LaoDong_27-12-2020	1976	13/01/2021
10_LaoDong_28-12-2020	10282	13/01/2021
10_LaoDong_30-12-2020	2020	13/01/2021
10_LaoDong_31-12-2020	21801	13/01/2021
10_ThanhNien_27-12-2020	7799	13/01/2021
10_ThanhNien_30-12-2020	4057	13/01/2021
10_ThanhNien_31-12-2020	7251	13/01/2021
10_TuoiTre_27-12-2020	1015	13/01/2021
10_TuoiTre_30-12-2020	26379	13/01/2021
10_TuoiTre_31-12-2020	13003	13/01/2021
11_26-12-2020	11520	13/01/2021
11_LaoDong_27-12-2020	1459	13/01/2021
11_LaoDong_28-12-2020	10426	13/01/2021
11_LaoDong_30-12-2020	1895	13/01/2021
11_LaoDong_31-12-2020	2098	13/01/2021
11_ThanhNien_27-12-2020	2409	13/01/2021
11_ThanhNien_30-12-2020	13492	13/01/2021

Remove Add



C:\Users\lambert28\Documents\NooJ\vi\Lexical_Analysis\VN_NP_Fonctions.dic

Dictionary contains 919 entries

Dictionnaires
Personnes et
organismes

thứ trưởng bộ LĐTBXH, NP+FONC
chủ tịch HĐND tỉnh, N+FONC
phó bí thư tỉnh ủy, N+FONC
phó chủ tịch UBND tỉnh, N+FONC
bí thư tu đoàn thanh niên, NP+FONC
bí thư huyện ủy, NP+FONC
phó chủ tịch UBND tỉnh, N+FONC
phó chủ tịch UBND tỉnh, N+FONC
chủ tịch HĐQT ngân hàng công thương VN, NP+FONC
phó chủ tịch UBND tỉnh, N+FONC
trưởng ban dân vận tỉnh ủy, NP+FONC
tỉnh trưởng ngãi, NP+FONC
nguyên hoàng anh, NP+FONC=bí thư tỉnh ủy cao bằng
chủ ngọc anh, NP+FONC=thứ trưởng bộ khoa học, NP+FONC=công nghệ
nguyên thủy anh, NP+FONC=phó chủ nhiệm các vấn đề xã hội của qh
trần tuần anh, NP+FONC=thứ trưởng bộ công thương
nguyên xuân anh, NP+FONC=bí thư thành ủy đã nằng

C:\Users\lambert28\Documents\NooJ\vi\Lexical_Analysis\VN_EN.dic

Dictionary contains 3710 entries

Cấm nạng mua sắm, N+EN+Organism+Press_Organ
Chuyên san Đồ Gia Dụng Việt Nam, N+EN+Organism+Press_Org
Daily Việt Nam - Tổng hợp và tìm kiếm tin tức, N+EN+Orga
Diễn đàn nhân tâm Đông đội, N+EN+Organism+Press_Organ
Đài tiếng nói Việt Nam, N+EN+Organism+Press_Organ
Đài truyền hình Việt Nam, N+EN+Organism+Press_Organ
Tạp chí Công nghệ thông tin, N+EN+Organism+Press_Organ
Le Courier du Vietnam, N+EN+Organism+Press_Organ
Nghề báo, N+EN+Organism+Press_Organ
Tạp chí Cộng sản, N+EN+Organism+Press_Organ
Tạp chí chuyên đề Internet của VDC, N+EN+Organism+Press_
Tạp chí Echip, N+EN+Organism+Press_Organ
Tạp chí Gia đình, N+EN+Organism+Press_Organ
Tạp chí Hàng Hải, N+EN+Organism+Press_Organ
Tạp chí Kế toán, N+EN+Organism+Press_Organ
Tạp chí Kiến trúc Nhà đẹp, N+EN+Organism+Press_Organ
Tạp chí Lao động, N+EN+Organism+Press_Organ
Tạp chí Ngàv Nav, N+EN+Organism+Press_Organ

```

---
aliases: []
tags: [Trần_Thị_Quốc_Khánh]
---

```

Concordance
NooJ

```

<?xml version="1.0" encoding="utf-8"?>
<Texte>
<Url>https://laodong.vn/xa-hoi/cong-bo-10-su-kien-y-te-va-phong-chong-dich-
covid-19-cua-viet-nam-2020-866655.ldo</Url> <Title>Công bố 10 sự kiện y tế và
phòng chống dịch COVID-19 của Việt Nam 2020</Title>
  <Date>None</Date>
  <Author>[Thùy Linh]</Author>
  <Content>Bộ Y tế vừa công bố 10 sự kiện y tế và phòng chống dịch bệnh COVID-19
của Việt Nam 2020.

```

1. Việt Nam thành công trong phòng chống dịch COVID-19

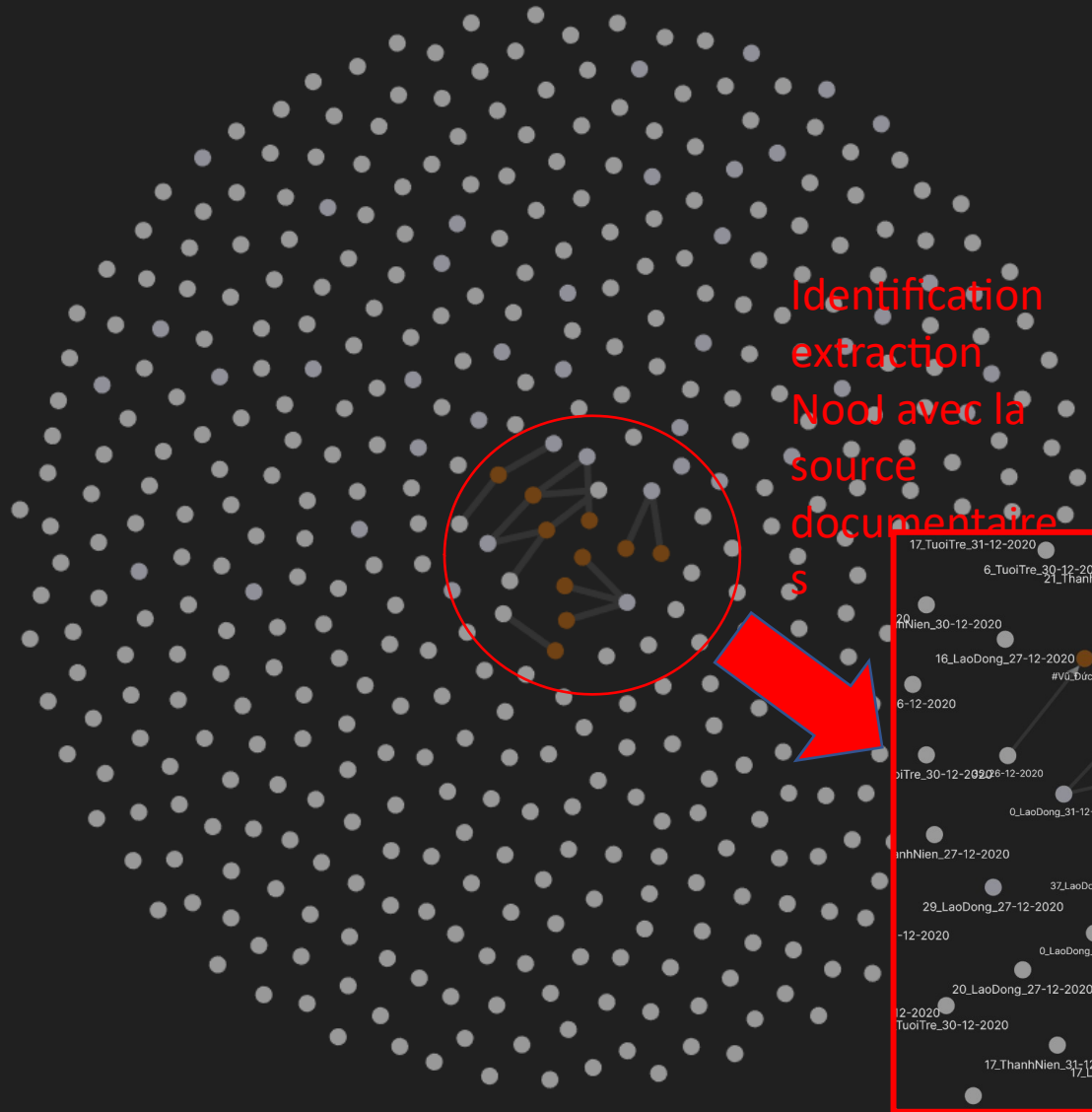
Việt Nam được cộng đồng quốc tế đánh giá là quốc gia thành công nhất trong phòng, chống dịch COVID-19, huy động được toàn bộ hệ thống chính trị, toàn Đảng, toàn dân, toàn quân tham gia phòng chống COVID-19 với chi phí thấp nhất.

Với chiến lược “ngăn chặn, phát hiện, cách ly, khoanh vùng, dập dịch và điều trị”, Việt Nam đã thực hiện được cách ly cho hơn 730.000 người; thực hiện xét nghiệm cho 1,7 triệu người; triển khai 1608 điểm chốt biên giới với gần 10.000 người bám biên từ Tết Canh Tý đến nay.

Bộ Y tế lần đầu tiên huy động lực lượng cán bộ y tế gần 300 người vào “mặt trận” Đà Nẵng triển khai công tác phòng, chống dịch tại vùng thành phố được điểm nóng của dịch bệnh.

Texte
source XML

Nguyễn_Phú_Trọng	3
Nguyễn_Xuân_Phúc	3
Vũ_Đức_Đam	2
Chu_Ngọc_Anh	1
Nguyễn_Hòa_Bình	1
Nguyễn_Thành_Phong	1
Nguyễn_Văn_Bình	1
Trần_Thanh_Mẫn	1
Trần_Thị_Quốc_Khánh	1
Võ_Văn_Thường	1





IX) Conclusion

- NOOJ compatible avec la DATAVIZ (perspectives/pistes : export CSV , API , PPlugin).
- NOOJ nécessite un certaine expertise toutefois on peut faire des extractions d'entités nommées sans être expert. (Conseillé de faire la semaine NOOJ).
- On peut traiter des langues peu dotées vs IA et modèle statistique.
- NOOJ peut avoir un usage technico-fonctionnel (perspectives : mode « Simplifié » et « analyste » vs mode « linguiste »)

NOOJ c'est le « QGIS » du traitement des données langagières.

