



HAL
open science

Le syntagme vietnamien

Philippe Lambert

► **To cite this version:**

Philippe Lambert. Le syntagme vietnamien. École thématique. Semaine NooJ à l'Inalco, INALCO, 2012. cel-00677777

HAL Id: cel-00677777

<https://cel.hal.science/cel-00677777>

Submitted on 9 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Le syntagme nominal vietnamien

Philippe Lambert

Institut Jean Lamour, Nancy – ASIÉS (Inalco)

philippe.lambert@ijl.nancy-universite.fr



PLAN

- 1. CONTEXTE**
- 2. PROBLEMATIQUE**
- 3. LE SYNTAGME NOMINAL EN VIETNAMEIEN**
- 4. LE SN DANS VIET4NooJ**
- 5. CONCLUSION**



Contexte

Problématique

Le SN viet.

Le SN et Viet4NooJ



Chinois classique
(Chữ Hán)
(1000 avant J.C env. à 1000 après J.C. env.)

Ecriture démotique vietnamienne
(Chữ Nôm)
(Du 10ème siècle après J.C. au début 20ème siècle après J.C.)

Langue nationale romanisée
(Quốc ngữ)
(Du 17ème siècle à maintenant)



La langue vietnamienne : écriture et phonétique

- 41 phonèmes
- 23 consonnes
- 13 voyelles simples
- 3 diphtongues
- 2 semi-voyelles
- 29 lettres

Aa Ăă Ââ Bb Cc Dd Đđ Ee Êê Gg Hh Ii Kk Ll Mm Nn Oo Ôô O'ơ
Pp Qq Rr Ss Tt Uu U'ư Vv Xx Yy

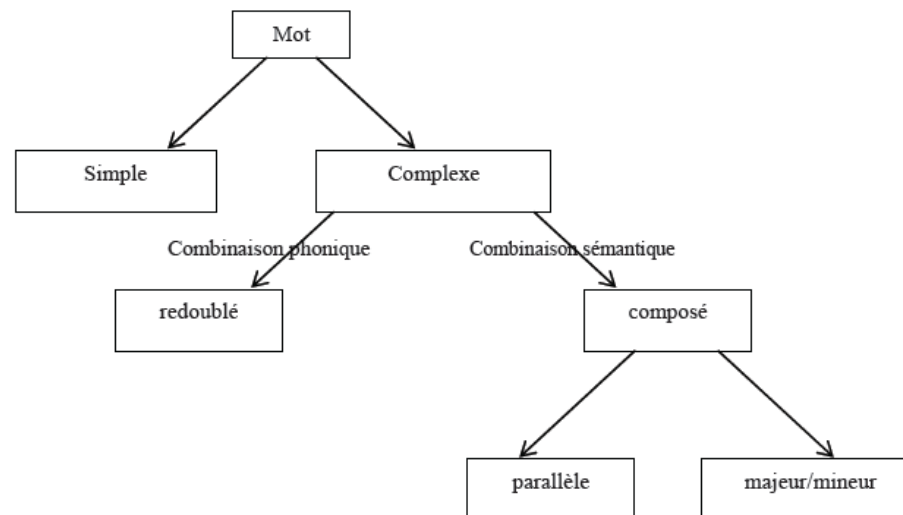
- 6 tons

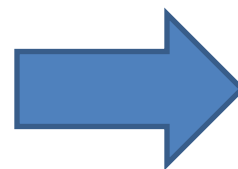


La langue vietnamienne :

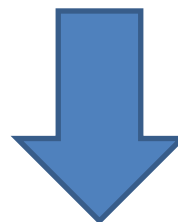
- Isolante
- Mots invariables
- Monosyllabique :

ex. : Anh đi chợ - 3 mots, 3 morphèmes, 3 syllabes





**Formalisation TAL
du vietnamien**



Problème :





Contexte

Problématique

Le SN viet.

Le SN et Viet4NooJ



Lê Văn Lý
Description de la langue
Logique combinatoire

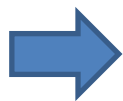
1948

Nguyễn Kim Thân, Nguyễn Tài Cẩn, Trần
Chút, Lê Xuân Thai, Lưu Văn Lang, Hoàng
Tuệ, Hoàng Phê
Approche sémantico-syntaxique
-> classes de mots

1983

Viet
4
NooJ

Dictionnaire (+40300 entrées)
**Grammaires (+ vingtaine : entités nommées, brevets,
etc.)**



Choix de catégories pour l'étiquetage



Le syntagme nominal vietnamien

Numéral	Classificateur	Substantif	Qualificatif	Démonstratif
- 2	- 1	0	+1	+2

Ba
Trois

Con
[class. animé]

Mèo
chat

Mun
noir

Kia
Là-bas

Ces trois chats noirs

Mot de totalité	Quantificateur	classificateur	Spécificatif	Nom	Complément (nom, adj., verbe)	Démonstratif
--------------------	----------------	----------------	--------------	-----	-------------------------------------	--------------

Tất cả
Totalité

năm
cinq

cái
[Class. Inanimé]

quyển
[spécif. livres]

sách
livre

cũ
ancien

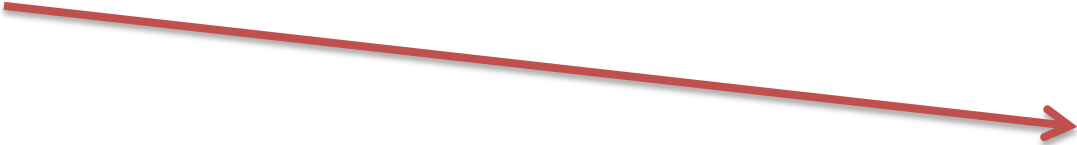
này
ces

Tous ces cinq vieux livres-ci

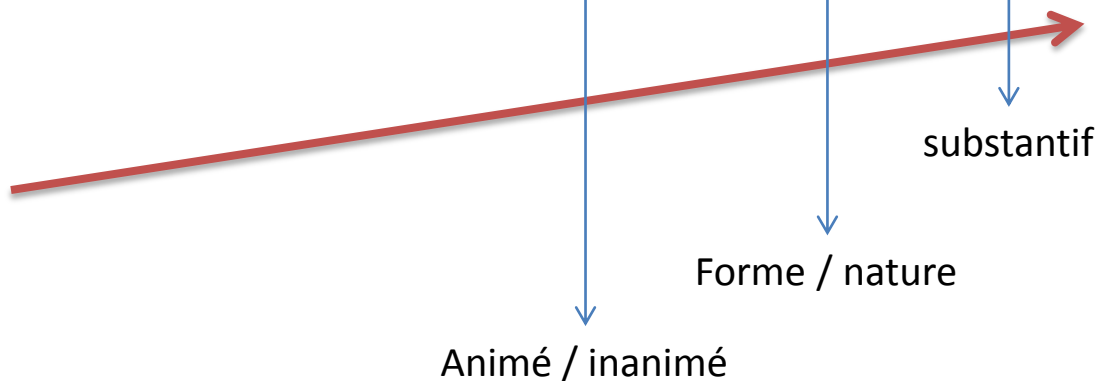


Le syntagme nominal vietnamien

Une approche cognitive ontologique



Mot de totalité	Quantificateur	classificateur	Spécificatif	Nom	Complément (nom, adj., verbe)	Démonstratif
-----------------	----------------	----------------	--------------	-----	-------------------------------	--------------





Le syntagme nominal vietnamien

Différentes écoles -> approche fonctionnaliste

Mot de totalité	Quantificateur	classificateur	Nom 1	Nom 2	Complément (nom, adj., verbe)	Démonstratif
--------------------	----------------	----------------	-------	-------	-------------------------------------	--------------

Chercheurs TAL – Hanoi – Loria/ Nancy :

Classificateur = spécificatif = « nom individuel »

(Lê Hồng Phương, *Elaboration d'un composant syntaxique à base de grammaires*, thèse de doctorat, Loria, Nancy, 2010)

OU

« **Nom Unit** » : réservé aux noms désignant une unité de mesure ou une entité.

o **natural** : Les classificateurs d'individus portent cette valeur, par ex. con = <classificateur d'animal>, bức = <classificateur d'un objet aplati>, ...

o **conventional** : unités de mesure conventionnelles ou scientifiques, par ex. mét = mètre, giờ = heure, nhúm = pincée, hào = centime, ...

o **collective** : Les classificateurs des collectifs portent cette valeur, par ex. nhóm = groupe, toán = troupeau, đôi = couple, chục = dizaine.

o **administrative** : unités administratives, par ex. tỉnh = province, xã = commune, ngành = division, môn = discipline

(Nguyễn Thị Minh Huyền, *Outils et ressources linguistiques pour l'alignement de textes multilingues français-vietnamiens*, thèse de doctorat, Loria, Nancy, 2006)



Contexte

Problématique

Le SN viet.

Le SN et Viet4NooJ

4. LE SN dans VIET4NooJ



Mot de totalité	Quantificateur	classificateur	Spécificatif	Nom	Complément (nom, adj., verbe)	Démonstratif
-----------------	----------------	----------------	--------------	-----	-------------------------------	--------------

Tất cả
Totalité

năm
cinq

cái

[Class. Inanimé]

quyển

[spécif. livres]

sách

livre

cũ

ancien

này

ces

Tous ces cinq vieux livres-ci



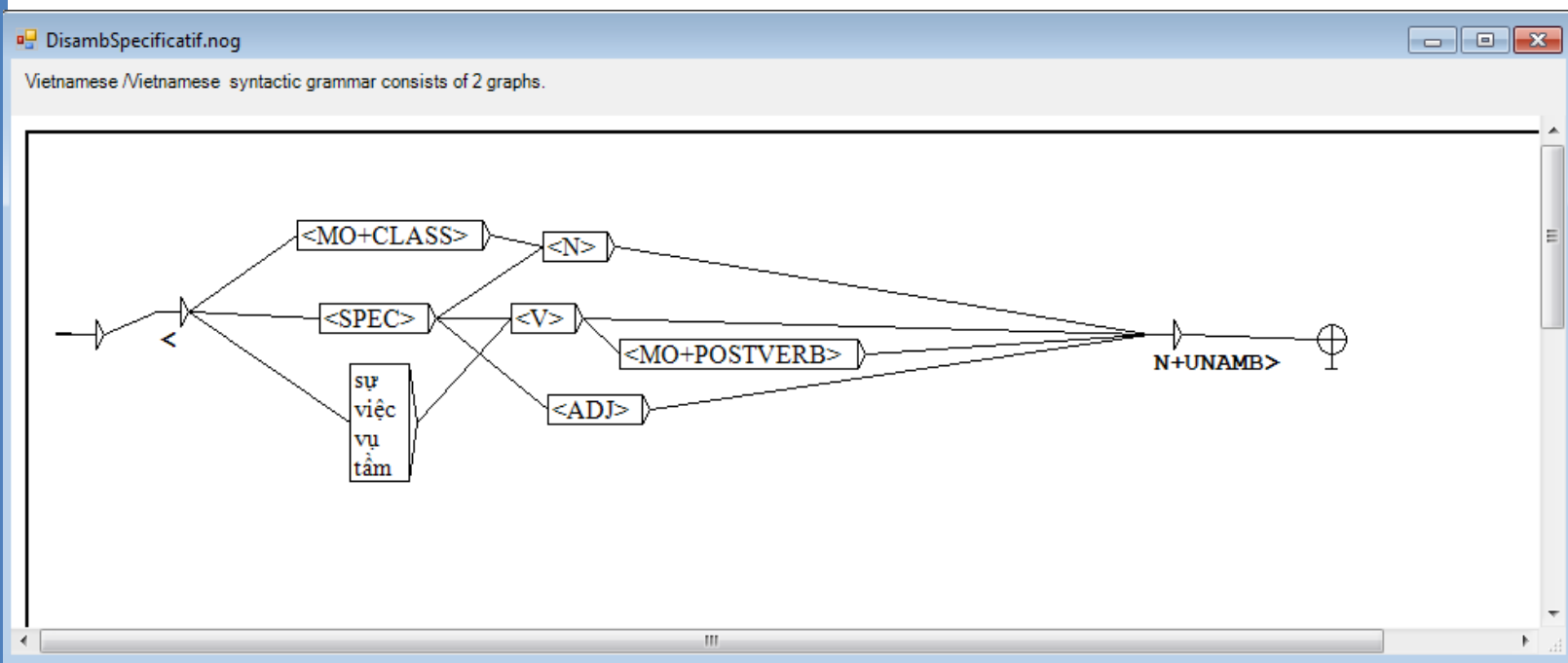
Un choix pragmatique et généraliste : l'étiquette <SPEC>

Dictionary contains 40250 entries

Entry	S-Lemma	Category	SynSem
bánh	bánh	SPEC	-
bao	bao	SPEC	-
bầu	bầu	SPEC	-
bây	bây	SPEC	-
bè	bè	SPEC	-
bẹ	bẹ	SPEC	-
bể	bể	SPEC	-
bình	bình	SPEC	-
bọn	bọn	SPEC	-
bồ	bồ	SPEC	-
bơ	bơ	SPEC	-
bụng	bụng	SPEC	-
buồng	buồng	SPEC	-
bữa	bữa	SPEC	-
cối	cối	SPEC	-
cơi	cơi	SPEC	-
cuộc	cuộc	SPEC	-
chăn	chăn	SPEC	-
chặng	chặng	SPEC	-
chăm	chăm	SPEC	-
châu	châu	SPEC	-
chiều	chiều	SPEC	-



Grammaire pour l'étiquetage des noms : <N>





Test sur un corpus :

Corpus DCSVN.noc

Locate a pattern in DCSVN.noc

acters

Concordance for Corpus DCSVN.noc

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

Before	Seq.	After
là do ảnh hưởng của thành tựu quan trọng trong bước phát triển, quy mô góp phần quan trọng vào vững, chất lượng, hiệu quả, sức cạnh tranh của chế kinh tế, chất lượng những điểm yếu cản trở cản trở sự phát triển.	cuộc khủng hoảng việc thực hiện nền kinh tế việc thực hiện sức cạnh tranh nền kinh tế nguồn nhân lực sự phát triển	tài chính, suy thoái kinh các mục tiêu, nhiệm vụ tăng lên. Đời sống của thắng lợi Chiến lược phát của nền kinh tế thấp thấp; chuyển dịch cơ cấu, kết cấu hạ tầng vẫn . Nền dân chủ xã hội và sức mạnh đại đoàn tài chính, suy thoái kinh của các thế lực thù ; nhưng trực tiếp và quyết , chỉ đạo của Đảng, quản độc lập, tự chủ, giữ chất lượng, hiệu quả tăng với tốc độ tăng trưởng trong và ngoài nước, sử để đẩy nhanh tốc độ . Ba là, phải coi trọng chặt chẽ giữa tăng trưởng , đồng bào ở vùng sâu dân chủ đi đôi với cao, giỏi về chuyên môn thuận cao, phát huy sức

Query 74/74



Contexte

Problématique

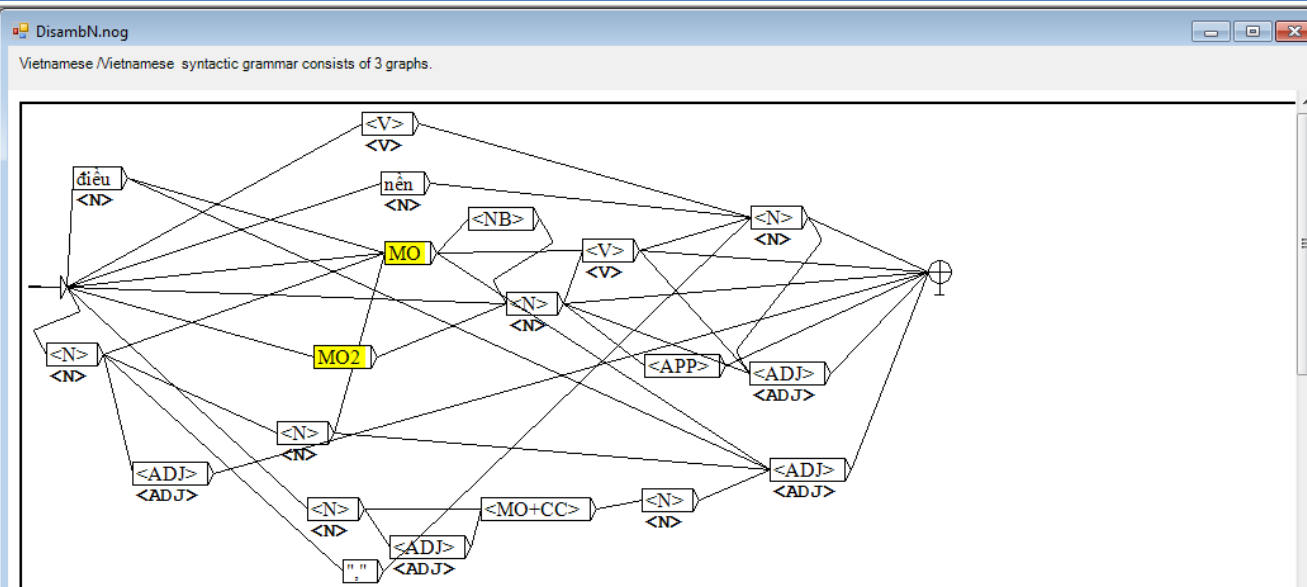
Le SN viet.

Le SN et Viet4NooJ

Structure of DisambN

Expand Collapse Refresh

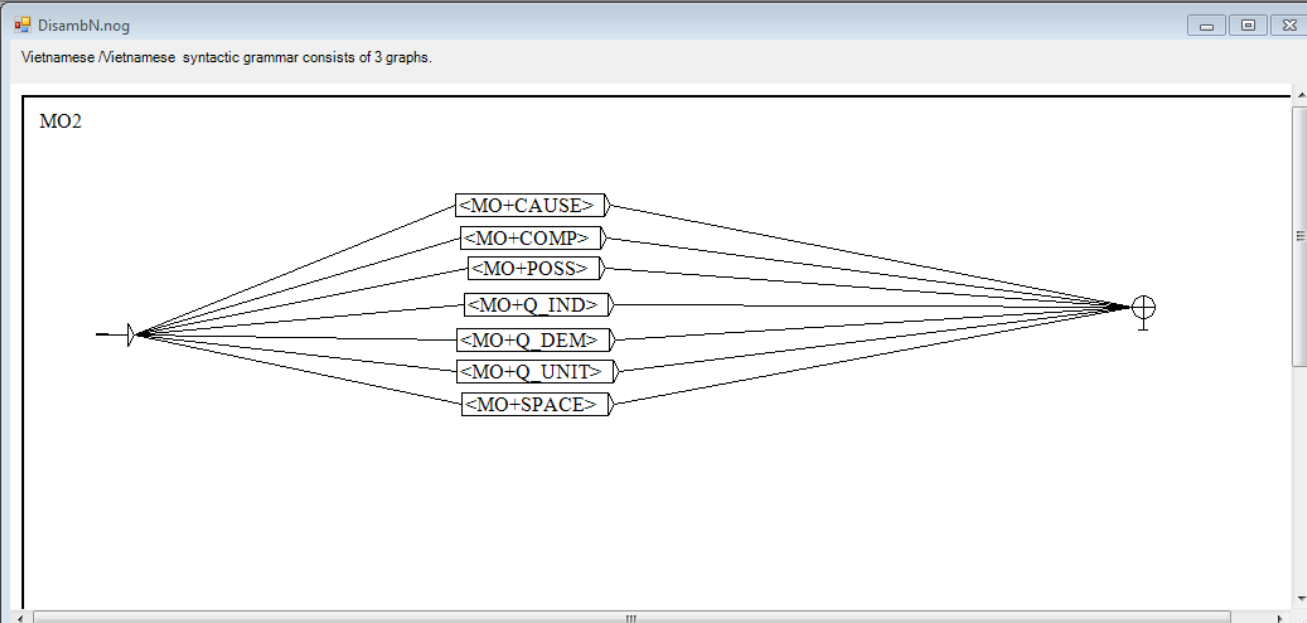
- Main
 - MO
 - MO2



Structure of DisambN

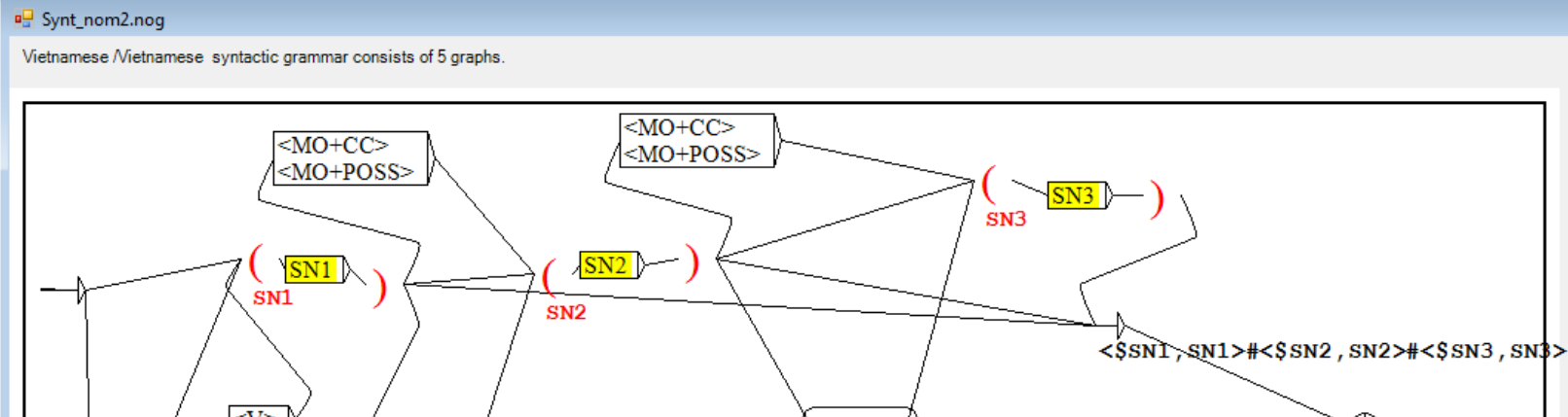
Expand Collapse Refresh

- Main
 - MO
 - MO2**





Les grammaires du syntagme nominal:



Concordance for Corpus Corpus DCSVN.noc

Reset Display: 0 characters before, and 0 after. Display: Matches Outputs

Text	Before	Seq.
NghiquestDaiHoi11.not		ngày 20 tháng 1 năm</N><NB><N><NB><N><ngày,SN1><20 tháng,SN2><1 năm,SN3>
NghiquestDaiHoi11.not		Nghi quyết Đại hội đại biểu toàn quốc</N><N><N><ADJ><Nghi quyết,SN1><Đại hội,SN2><đại biểu toàn quốc,SN3>
NghiquestDaiHoi11.not		ĐẠI HỘI ĐẠI BIỂU TOÀN QUỐC</N><N><N><ĐẠI HỘI,SN1><ĐẠI BIỂU,SN2><TOÀN QUỐC,SN3>
NghiquestDaiHoi11.not		Đại hội đại biểu toàn quốc</N><N><N><Đại hội,SN1><đại biểu,SN2><toàn quốc,SN3>
NghiquestDaiHoi11.not		các văn kiện do Ban Chấp hành Trung ương Đảng khoá</MO><N><MO><N><ADJ><N><các văn kiện,SN1><Ban Chấp hành Trung ương Đảng,SN2><khoá,SN3>
NghiquestDaiHoi11.not		chính trị của Ban Chấp hành Trung ương Đảng khoá</N><MO><N><ADJ><N><chính trị,SN1><Ban Chấp hành Trung ương Đảng,SN2><khoá,SN3>
NghiquestDaiHoi11.not		chính trị của Ban Chấp hành Trung ương Đảng khoá</N><N><ADJ><N><chính trị,SN1><Ban Chấp hành Trung ương Đảng,SN2><khoá,SN3>
NghiquestDaiHoi11.not		ảnh hưởng của cuộc khủng hoảng tài chính</N><MO><N><N><ảnh hưởng,SN1><cuộc khủng hoảng,SN2><tài chính,SN3>
NghiquestDaiHoi11.not		ảnh hưởng của cuộc khủng hoảng tài chính</N><N><N><ảnh hưởng,SN1><cuộc khủng hoảng,SN2><tài chính,SN3>
NghiquestDaiHoi11.not		toàn quân ta</N><N><APP><toàn,SN1><quân,SN2><ta,SN3>
NghiquestDaiHoi11.not		những thành tựu quan trọng trong việc thực hiện các mục tiêu</MO><N><ADJ><MO><N><MO><N><những thành tựu quan trọng,SN1><việc thực hiện,SN2><các mục tiêu,SN3>
NghiquestDaiHoi11.not		những thành tựu quan trọng trong việc thực hiện các mục tiêu</MO><N><ADJ><MO><N><MO><N><những thành tựu quan trọng,SN1><việc thực hiện,SN2><mục tiêu,SN3>
NghiquestDaiHoi11.not		kết quả trước những diễn biến phức tạp của kinh tế</N><MO><MO><N><ADJ><MO><N><kết quả,SN1><những diễn biến phức tạp,SN2><kinh tế,SN3>
NghiquestDaiHoi11.not		kết quả trước những diễn biến phức tạp của kinh tế</N><MO><MO><N><ADJ><N><kết quả,SN1><những diễn biến phức tạp,SN2><kinh tế,SN3>
NghiquestDaiHoi11.not		kết quả trước những diễn biến phức tạp của kinh tế</N><MO><MO><N><ADJ><N><kết quả,SN1><diễn biến phức tạp,SN2><kinh tế,SN3>
NghiquestDaiHoi11.not		kết quả trước những diễn biến phức tạp của kinh tế</N><MO><MO><N><ADJ><N><kết quả,SN1><diễn biến phức tạp,SN2><kinh tế,SN3>
NghiquestDaiHoi11.not		uy tín quốc tế của nước ta</N><ADJ><MO><N><APP><uy tín quốc tế,SN1><nước,SN2><ta,SN3>
NghiquestDaiHoi11.not		uy tín quốc tế của nước ta</N><ADJ><N><APP><uy tín quốc tế,SN1><nước,SN2><ta,SN3>
NghiquestDaiHoi11.not		khởi đại đoàn kết toàn dân tộc</N><N><ADJ><N><khởi,SN1><đại đoàn kết toàn,SN2><dân tộc,SN3>
NghiquestDaiHoi11.not		khởi đại đoàn kết toàn dân tộc</N><N><MO><N><khởi,SN1><đại đoàn kết,SN2><dân tộc,SN3>



Extraction des variables <SN> à partir du corpus test :

Concordance for Corpus Corpus DCSVN.noc

Reset Display: 0 characters before, and 0 after. Display: Matches Outputs

Text	Before	Seq.
NghiquestDaiHoi11.not		ngày 20 tháng 1 năm/<N><NB><N><NB><N><ngày,SN1><20 tháng,SN2><1 năm,SN3>
NghiquestDaiHoi11.not		Nghi quyết Đại hội đại biểu toàn quốc/<N><N><N><ADJ><Nghi quyết,SN1><Đại hội,SN2><đại biểu toàn quốc,SN3>
NghiquestDaiHoi11.not		ĐẠI HỘI ĐẠI BIỂU TOÀN QUỐC/<N><N><N><ĐẠI HỘI,SN1><ĐẠI BIỂU,SN2><TOÀN QUỐC,SN3>
NghiquestDaiHoi11.not		Đại hội đại biểu toàn quốc/<N><N><N><Đại hội,SN1><đại biểu,SN2><toàn quốc,SN3>
NghiquestDaiHoi11.not		các văn kiện do Ban Chấp hành Trung ương Đảng khoá/<MO><N><MO><N><ADJ><N><các văn kiện,SN1><Ban Chấp hành Trung ương Đảng,SN2><khóa,SN3>
NghiquestDaiHoi11.not		chính trị của Ban Chấp hành Trung ương Đảng khoá/<N><MO><N><ADJ><N><chính trị,SN1><Ban Chấp hành Trung ương Đảng,SN2><khóa,SN3>
NghiquestDaiHoi11.not		chính trị của Ban Chấp hành Trung ương Đảng khoá/<N><N><ADJ><N><chính trị,SN1><Ban Chấp hành Trung ương Đảng,SN2><khóa,SN3>
NghiquestDaiHoi11.not		ảnh hưởng của cuộc khủng hoảng tài chính/<N><MO><N><N><ảnh hưởng,SN1><cuộc khủng hoảng,SN2><tài chính,SN3>
NghiquestDaiHoi11.not		ảnh hưởng của cuộc khủng hoảng tài chính/<N><N><N><ảnh hưởng,SN1><cuộc khủng hoảng,SN2><tài chính,SN3>
NghiquestDaiHoi11.not		toàn quân ta/<N><N><APP><toàn,SN1><quân,SN2><ta,SN3>
NghiquestDaiHoi11.not		những thành tựu quan trọng trong việc thực hiện các mục tiêu/<MO><N><ADJ><MO><N><MO><N><những thành tựu quan trọng,SN1><việc thực hiện,SN2><các mục tiêu,SN3>
NghiquestDaiHoi11.not		những thành tựu quan trọng trong việc thực hiện các mục tiêu/<MO><N><ADJ><MO><N><MO><N><những thành tựu quan trọng,SN1><việc thực hiện,SN2><mục tiêu,SN3>
NghiquestDaiHoi11.not		kết quả trước những diễn biến phức tạp của kinh tế/<N><MO><MO><N><ADJ><N><kết quả,SN1><những diễn biến phức tạp,SN2><kinh tế,SN3>
NghiquestDaiHoi11.not		kết quả trước những diễn biến phức tạp của kinh tế/<N><MO><MO><N><ADJ><MO><N><kết quả,SN1><diễn biến phức tạp,SN2><kinh tế,SN3>
NghiquestDaiHoi11.not		kết quả trước những diễn biến phức tạp của kinh tế/<N><MO><MO><N><ADJ><N><kết quả,SN1><diễn biến phức tạp,SN2><kinh tế,SN3>
NghiquestDaiHoi11.not		uy tín quốc tế của nước ta/<N><ADJ><MO><N><APP><uy tín quốc tế,SN1><nước,SN2><ta,SN3>
NghiquestDaiHoi11.not		uy tín quốc tế của nước ta/<N><ADJ><N><APP><uy tín quốc tế,SN1><nước,SN2><ta,SN3>
NghiquestDaiHoi11.not		khối đại đoàn kết toàn dân tộc/<N><N><ADJ><N><khối,SN1><đại đoàn kết toàn,SN2><dân tộc,SN3>
NghiquestDaiHoi11.not		khối đại đoàn kết toàn dân tộc/<N><N><MO><N><khối,SN1><đại đoàn kết,SN2><dân tộc,SN3>

Query 380/380



Éléments bibliographiques

- [1] B. Q. Ho, “Vers une indexation structurée basée sur des syntagmes nominaux (impact sur un SRI en vietnamien et la RI multilingue),” 2004.
- [2] P. T. Nguyen, X. L. Vu, T. M. H. Nguyen, V. H. Nguyen, and H. P. Le, “Building a large syntactically-annotated corpus of Vietnamese,” in *Proceedings of the Third Linguistic Annotation Workshop*, 2009, pp. 182–185.
- [3] T. M. H. Nguyen, “Outils et ressources linguistiques pour l’alignement de textes multilingues français-vietnamiens,” 2006.
- [4] H. P. Le, T. M. H. Nguyen, A. Roussanaly, and T. V. Ho, “A hybrid approach to word segmentation of Vietnamese texts,” 2008.



Contexte

Problématique

Le SN viet.

Le SN et Viet4NooJ

Merci !
Xin cảm ơn !