

Komotini, Greece
May 27-29, 2010

Université Paul-Valéry



Montpellier III

Arts · Lettres · Langues
Sciences humaines & sociales



N o o J
2010

Automatic processing of temporality for VIET4 Nooj:

A demonstration of temporal frames detection

Nicolas BOFFO

Praxiling, UMR 5267 CNRS, Montpellier 3 University
MICA UMI 2954 CNRS, Hanoi University Technology

Océane HO DINH
INALCO



Plan

I. Objectives

II. Linguistic positions and orientations

III. Dictionary of aspectual and temporal terms

IV. Temporal frames: Definition

V. Dictionaries of temporal framing marks under NOOJ.

VI. Temporal frames detection: How is it done?

VII. Module of temporal framing detection in NOOJ

VIII. Demonstration of NOOJ

IX. Conclusion: Results, difficulties/needs, perspectives

I. Objectives

Objectives

- Delimit the temporal frames
- Determine the temporal values of the temporal frames
- Determine the temporal values of the corpus
- Locate the events in the time space

II. Linguistic positions and orientations

Our positions and orientation in linguistics: Praxematic

- Praxematic = production of meaning (actualization, polysemy, monosemy),
The main idea praxematic is that humans base their linguistic representations on their praxis (experience).

How we can combine the praxematic approach with the formal approach?

- It is our first difficulties.

- To resolve it:

- > Actually we implement regularly new semes according to the world knowledges.**
- > When we need we create a new contextual rules.**

Our positions and orientations: Temporality

- “SdT”(Semantic of Temporality) of L. Gosselin (cotext and context) > 3 intervals from Reichenbach points:

Reichenbach utilise trois points sur un axe temporel : S le moment de l'énonciation (*point of speech*), E le moment du procès (*point of the event*), et R le moment de référence (*point of reference*) qui correspond à un point temporel *d'observation* du procès. (p. 27, Thèse de Cédric Person: Traitement automatique de la temporalité du récit: Implémentation du modèle linguistique SdT)

=> SdT does a good job of exploiting aspectual and temporal relations but we noticed some difficulties in automatic processing of temporality because of the 3 intervals (point of view shared with S. Shwer from LIPN of University Paris 13)

- J. Bres & G.J. Barcelo: *Les temps de l'indicatif en français*
- S-Language of S. Schwer (Formal language which can treat the aspectuo-temporal relations)
- Temporal framing marks give the temporal value (cf. *supra*)

■ In Vietnamese there is time (≠ tense), aspect and modality.

■ In Vietnamese there are marks of TAM.

**Đã, đang and sẽ are not temporal marks but aspectual: 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. (If you run a red light, you are penalized.)

> **sẽ = action is after.**

Time and aspect categories

Time categories

Past

Present

Future

Neutral

Allen's relations

Aspect categories

Accomplished (đã)

In progress (đang)

Iterative (V+lại)

Terminative (xong, hết, thôi, etc.)

Imminent (sắp)

...

III. Dictionary of aspectual and temporal terms

Dictionary of aspectual and temporal terms

The aspectual and temporal marks (ATM) are classified in different categories according to their specification and the aspectual and temporal values that they can have in cotext and context.

Dictionary of ATM help us to tag the temporal and aspectual expressions (dictionary is currently under construction).

In our presentation we will present you the first step of automatic processing of temporality: the detection of the temporal framing marks which give the time value.

IV. Temporal frames: definition

Temporal frames: definition

- These are “the frames of the discourse” (Charolles, 1997).
 - Charolles identifies segments called “speech frames” which are homogeneous compared with a semantic criterion specified by an expression that is in the initial position: **Initiator of frame** (IF).
 - The IF are like marks of indexation which make it possible to regroup the prepositional contents in homogeneous blocks.
- => That is why it is very interesting to detect the IF.

Framing relations

Framing relations are supported

- by Adverbials (PP, Adv, SN, Subordinate Sentences): adjuncts, parentheticals
- In Adverbials: we have regular expression:
 - Date : sinh ngày 19 tháng 5 năm 1890
 - Period : Từ năm 1912 đến năm 1917

Source : M. Charolles, *Coherence and Cohesion Relations : Connexion and Framing*, p18, <http://www.lattice.cnrs.fr/Michel-Charolles>,152

Temporal Framing relations

Năm ngoái (IF)

Từ 12/09 đến 24/10 (IF)

Một ngày (IF)

Từ 19/09 đến 09/11 (IF)

V. Dictionaries of temporal framing marks

Dictionaries of temporal framing marks

The temporal framing marks are classified by time categories:

- past
- present
- future
- neutral

VI. Temporal frames detection: how is it done?

Temporal frames detection: How is it done?

The procedure of detection of temporal frames:

→ Detect an IF with the dictionaries implemented in NOOJ.

→ Find the end of the TF:

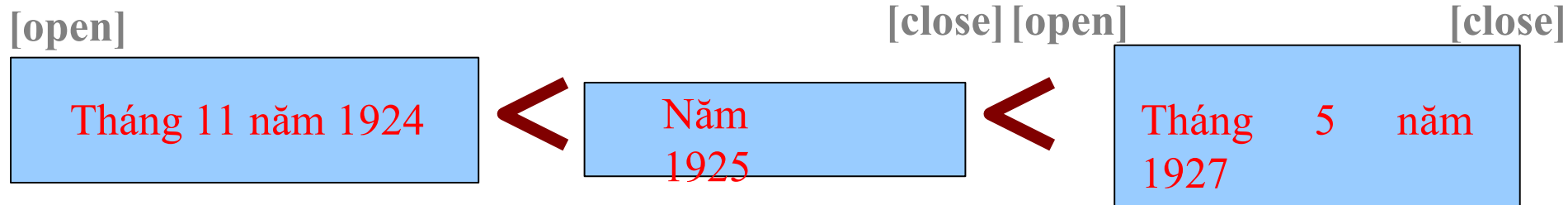
To know when the TF will be closed we have 3 indications:

1. A new occurrence of an IF;
2. A change of the time (difficult to find in Vietnamese because there is no inflection for the tenses);
3. A change of paragraph.

[open] abc... [close]

[open]Tháng 11 năm 1924, Nguyễn Ái Quốc về Quảng Châu (Trung Quốc)
Năm 1925, Người thành lập Hội Việt Nam Cách mạng Thanh niên[close]

[open]Tháng 5 năm 1927, Nguyễn Ái Quốc rời Quảng Châu đi Mátxcova (Liên Xô) ,sau đó đi Berlin (Đức), đi Bruxell (Bỉ) tham dự phiên họp mở rộng của Đại hội đồng Liên đoàn chống chiến tranh đế quốc, sau đó đi Ý và từ đây về Châu Á.[close]

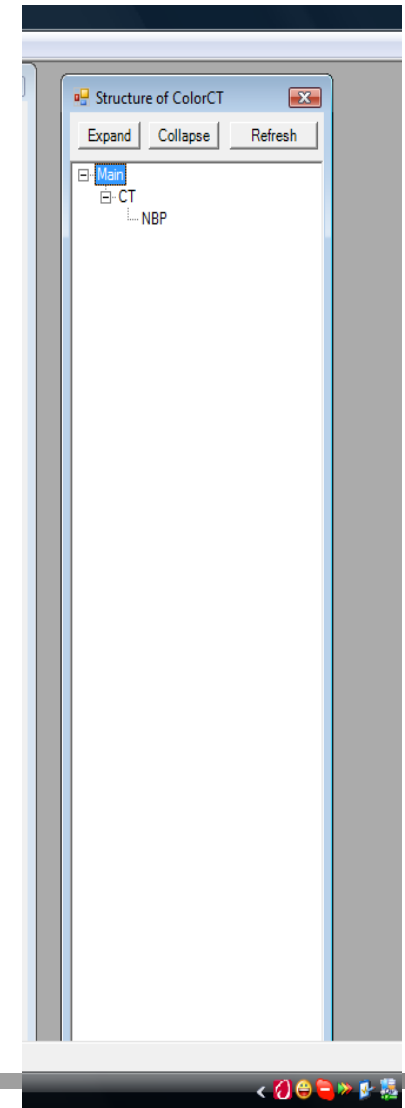
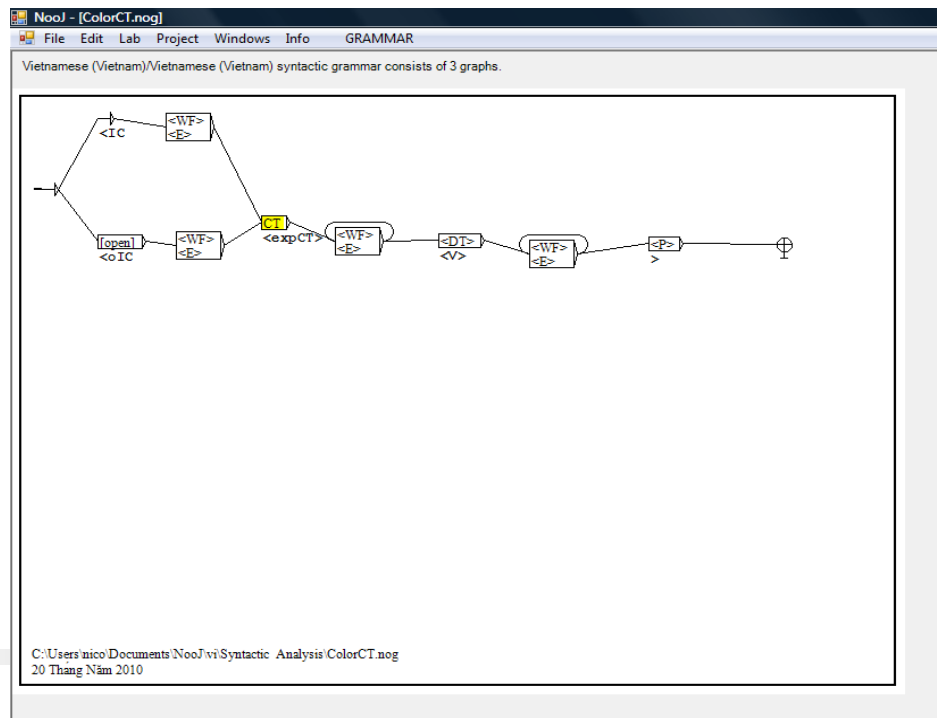


VII. Module of temporal frames detection on NOOJ

Module of temporal frames detection in NOOJ

The module of temporal frames detection in NOOJ consists of:

- one main graph (textual organization graph)
- sub-graphs (lexico-syntactic graphs)



How works the module of temporal frames detection ?

Actually the module detect most of the Initiators of Frame (IF) and part of the temporal frames with the punctuation and the verb's dictionary.

It's working:

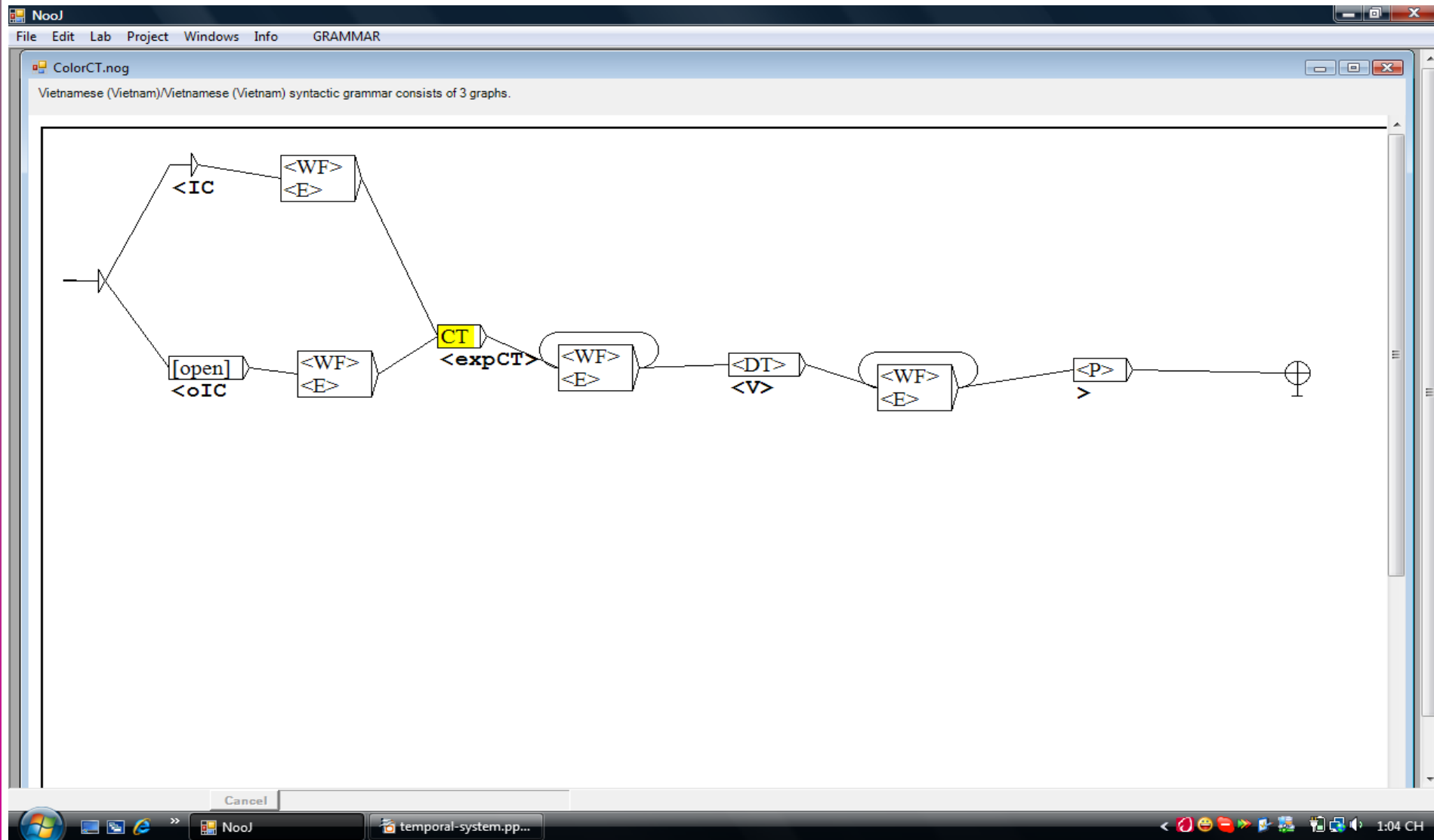
- by regular expressions which come from the work of formalization of temporality in Vietnamese (thesis research of Nicolas Boffo, Praxiling/MICA).
- with dictionnaires of temporal marks which are IF which are implemented in regular expressions.

The module tag correctly by time value most of IF.

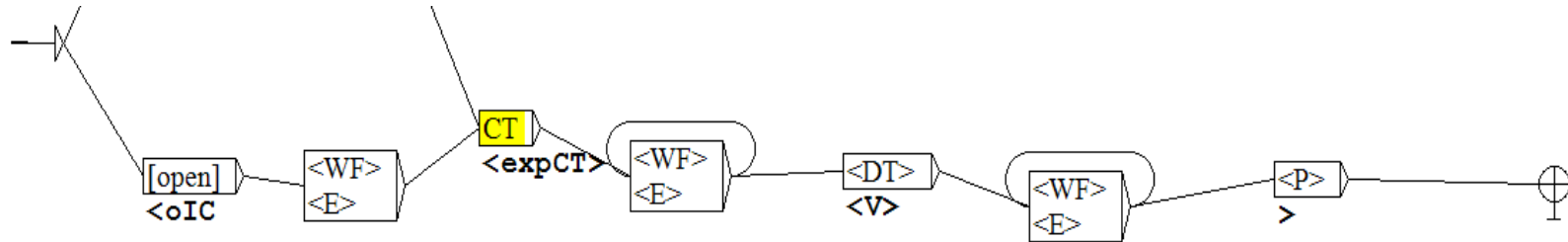
Unfortunately the module can't calculate the temporal relations (it is linked to the limit of NOOJ) and is limited to treat the dates.

Main graph: textual organization graph

Main graph: textual organisation graph

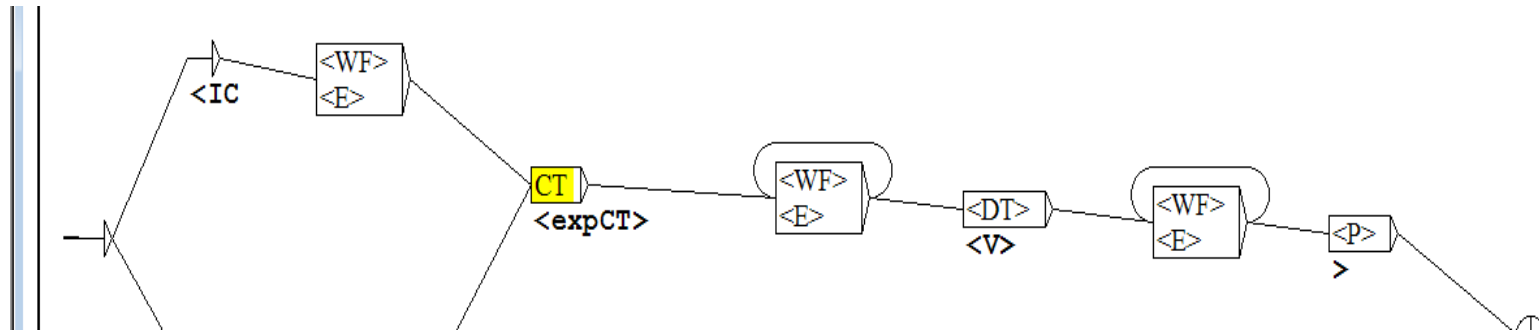


Main graph: IF at “[open]” position



Here we have the regular expression of the IF at “[open]” position. So we can detect and tag the IF in the beginning of paragraph by the Tag <oIC>. This will help us to know the change of temporal frame according to the 3 indications of the close of temporal frames.

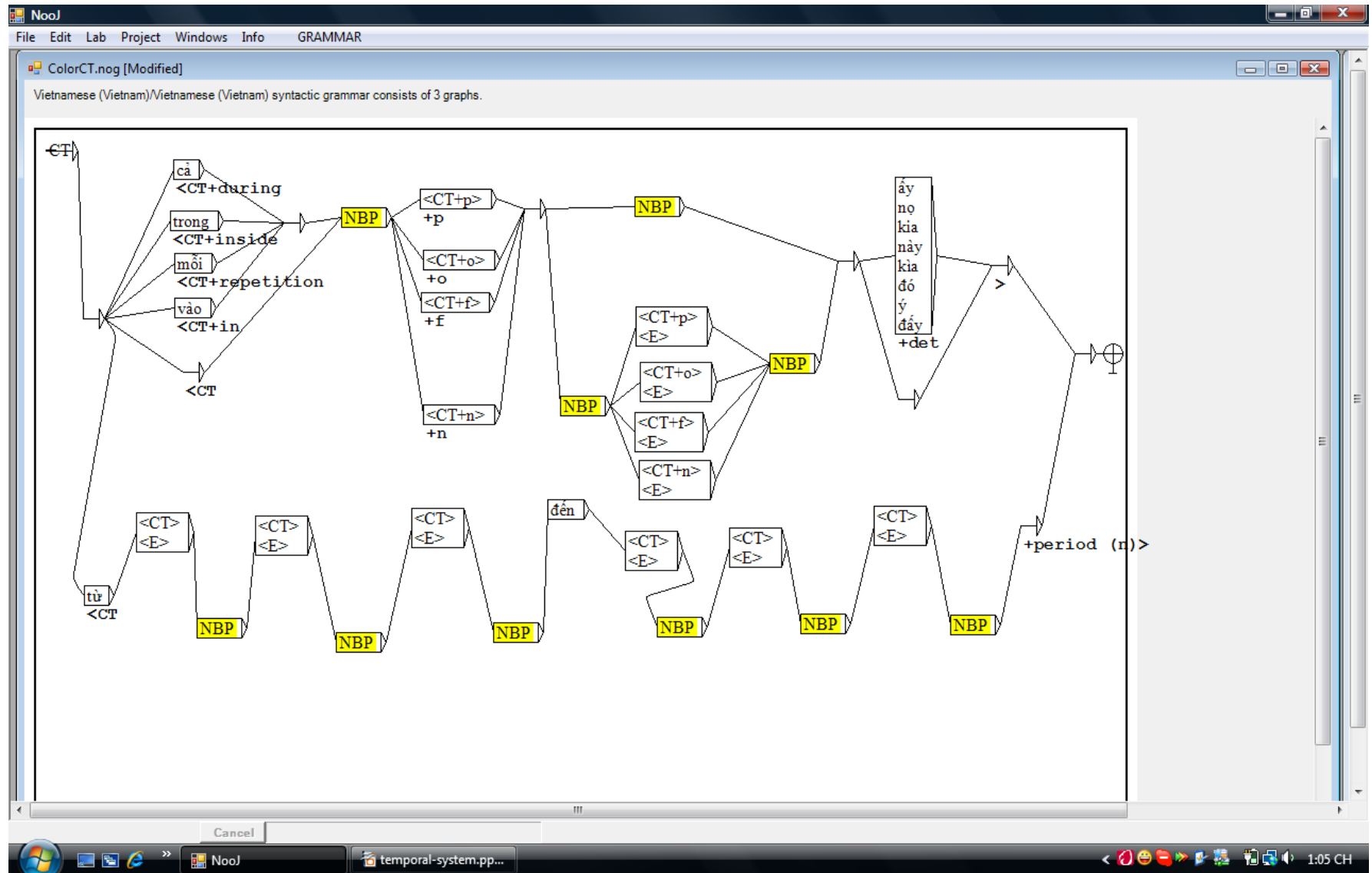
Main graph: IF inside the paragraph



This is the regular expression of all IF inside the paragraph.

Sub-graphs: lexico-syntactic graphs

Subgraph of IF



Subgraph of IF

This sub-graph is the combination of the most of regular expressions of IF that we can find in vietnamese language.

Exemples:

Mùa xuân (CT)/ năm (CT)/ 1975 (NB)

Spring/ year/number

Tuần(CT) này (det)

Week/this

Vào(CT+during) /năm(CT)/1890(number=

In/year/number

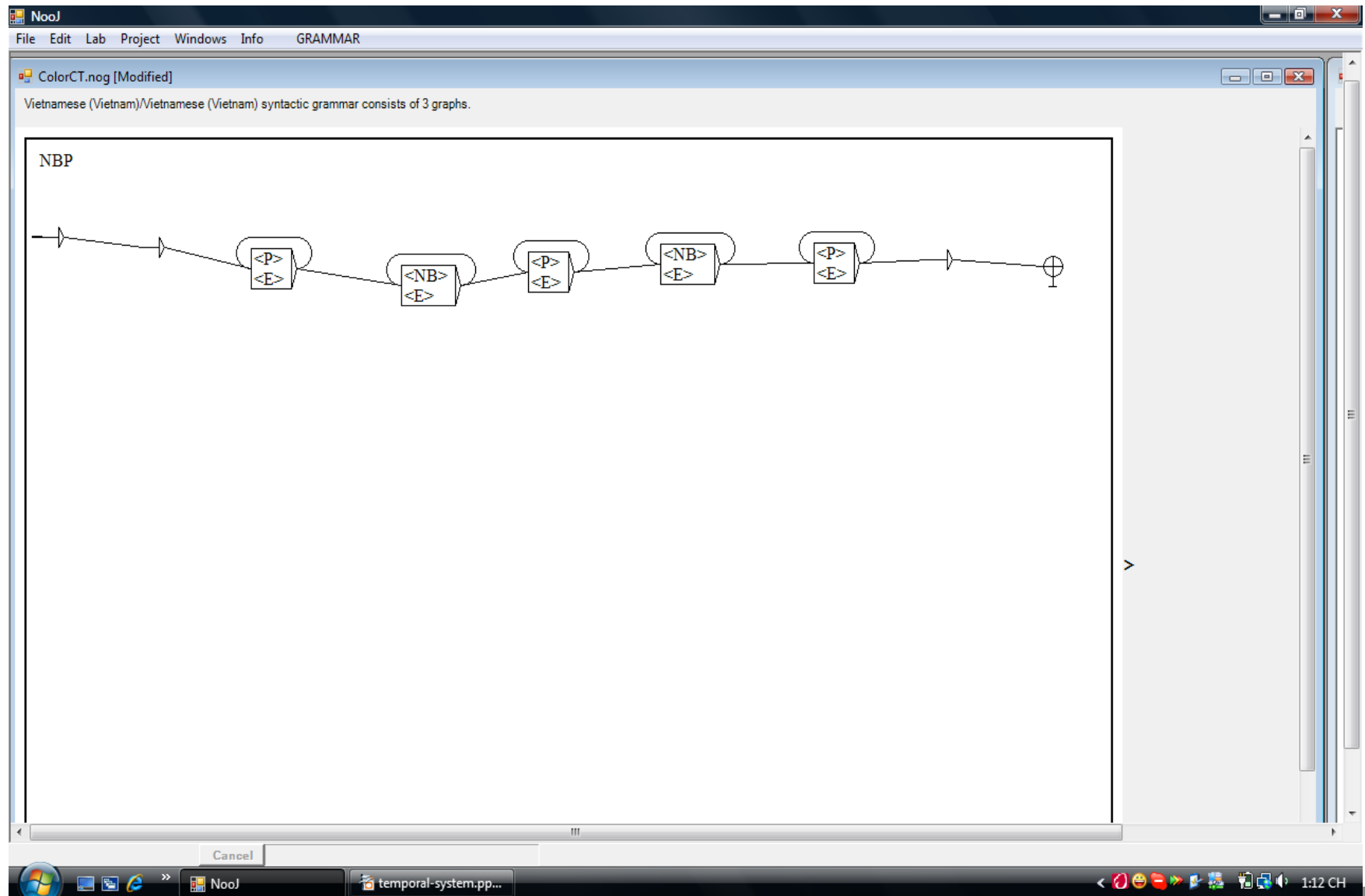
sinh ngày/ 19 tháng 5 năm 1890

Birth day/ nb/month/nb/year/nb

Từ /năm/ 1912/ đến/ năm /1917

from/year/nb/to/year/nb

Sub-graph of regular expressions for the numbers of dates



VIII. Demonstration of NOOJ

(Look the demonstration)

Constitution of corpus

- **Extraction from online newspapers.**
- **Preprocessing of HTML balise with script in Perl.**
- **Normalization**
- **Importation of texts in NOOJ**

IX. Conclusion: Results, difficulties/needs, perspectives

Results

- **We can see the temporal frames.**
- **We can give one part of temporal value.**
- **Better than the tools already developed by NOOJ project.**
- **Performance: we still calculating the recall and the precision because we need a good reference.**

Difficulties/needs

- **Calculate the chronology: we have to create a system for this.**
- **NOOJ can't treat the aspectuo-temporal relation.**
- **Need to implement some aspectuo-temporal rules**
- **Calculate the performance: which tools for temporal system do we have to use when we have a big corpus?**

Perspectives

- Find the temporal value.
- Increase the tools of automatic understanding.
- Detect the events.
- Find the position of the events in the temporal space.
- Predict the events.