



**HAL**  
open science

# Linguistics, applied research and NLP: using NooJ in a technical-operational context. Case-study, analysis and perspectives

Nicolas Boffo, Philippe Lambert

## ► To cite this version:

Nicolas Boffo, Philippe Lambert. Linguistics, applied research and NLP: using NooJ in a technical-operational context. Case-study, analysis and perspectives. 15 th INTERNATIONAL CONFERENCE NOOJ 2021, Jun 2021, Besançon (virtual conference), France. hal-03270741

**HAL Id: hal-03270741**

**<https://hal.science/hal-03270741v1>**

Submitted on 7 Dec 2022

**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.

# Linguistics, applied research and NLP: using NooJ in a technico-operational context. Case-study, analysis and perspectives

Nicolas Boffo<sup>1</sup> and Philippe Lambert<sup>2</sup>

<sup>1</sup> NLP Group, CIMI, France

<sup>2</sup> University of Lorraine, 34 Cours Léopold, 54000 NANCY, France

nicolas\_boffo@yahoo.com

philippe.lambert@univ-lorraine.fr

**Abstract.** NooJ is especially used to formalize languages and linguistic usages for research and didactic purposes, but also to be integrated into information processing technologies such as monitoring systems, semantic search engines, etc. Furthermore, NooJ can also be considered as a major system component in a very functional way or operational context. Nowadays, most of the information systems connected to economic intelligence must technically respond to the three important dimensions: reactivity, adaptability and interoperability. These constraints set up the background of our operational approach. The main purpose of our paper is to demonstrate the potential of NooJ not only for fundamental research in linguistics but also for operational objectives. Our paper will firstly present the linguistic work carried out on a poorly endowed language like Vietnamese for which we build up specific linguistics resources. Then, our study will present the first results of a visual representation of the information extracted by NooJ within a technico-functional uses: dataviz, spatial visualization of events with free and open source Geographic Information System QGIS and timeline visualization. Finally, we will analyze the uses of NOOJ in a technico-operational context by pointing out its weaknesses and strengths suggest innovative ways to improve NooJ tool.

**Keywords:** NLP, NOOJ, functional, operational, technico-operational, technico-functional, reactivity, adaptability, interoperability, dataviz, spatio-temporal visualization of events, QGIS, poorly endowed language, Vietnamese language.

## 1 Introduction

In "How the DEM and LVF can serve Digital Humanities" [20] ", the limits of certain statistical analysis tools such as Sketch Engine<sup>1</sup> are pointed out.

Indeed, the lexical-semantic fields calculated by this type of software "are not reliable" and "one should absolutely avoid selecting them when one wants to analyze the

---

1 Cf. Kilgariff (2004)

archives of a daily newspaper to estimate the evolution of the morale of the French or to find indices of depression in psychological interviews, at the risk of obtaining totally random results". [20]

Moreover, Max Silberztein concluded : "We hope that social scientists who need to explore and analyze their corpus will appreciate the qualitative gain obtained when one analyzes a text from its linguistic units (concepts, entities, predicates and relations) rather than from its simple graphic forms." [20]

Our approach will be less dichotomous : Expert Systems vs Stochastic Systems.

Indeed, this opposition between the expert system and stochastic system is even stronger nowadays by the fact that Artificial Intelligence (AI) with Deep Learning models is becoming is gaining momentum.

By the way, in this regard, there was a turning point in 2018 in the field of NLP with the arrival of the BERT model (Bidirectional Encoder Representations from Transformers) from Google, and more recently with the ERNIE<sup>2</sup> model from the Chinese tech giant Baidu which dethroned the two American giants Google and Microsoft in the ranking<sup>3</sup> of General Language Understanding Evaluation, otherwise known as GLUE<sup>4</sup>.

Despite these advances, is this the end of expert systems and symbolic AI?

No, far from it.

We are going to show that expert systems such as NOOJ have their place, especially in an area that is not economically profitable for NLP companies and that annoys the IT System Department of organizations because it does not meet the strict rules of IT project management.

We will call this zone : the techno-operational and/or functional zone.

This area includes the following activities: discourses analysis, KMS, prospecting, investigation,..., all important activities for an organization where the NOOJ engine will have a real added value.

We will then ask the question to know if NOOJ is accessible for users working on language data with a need for dataviz and mapping in technico-operational context.

Therefore, our article will be structured as follows:

---

2 <https://github.com/PaddlePaddle/ERNIE>

3 <https://gluebenchmark.com/leaderboard>

4 <https://gluebenchmark.com/>

2. Define and conceptualize the technico-operational (or functional) approach.
3. Case-study : NOOJ in a technico-operational context
  - 3.1. Spatio-temporal treatment in a technico-operational/functional approach : the case of the Vietnamese press.
  - 3.2. Knowledge Management System (KMS)
  - 3.3. Applying NooJ Linguistics Analysis throughout PKM frameworks
4. Conclusion
5. Perspectives

## 2 Define and conceptualize the technico-operational/functional approach

Initially, we find the expression technico-operational within the armies which have the vocation to carry out operational missions.

The first approach is :

Technico-operational in the context of experimentation/simulation :

Indeed, we find first of all the "Technico-operational foresight studies (PTO), which provide elements of appreciation for the expression of future needs..., technico-operational studies (ETO), which are used to optimize, for future systems or within the framework of evolutions of existing systems, the operational characteristics, technical specifications as well as the associated costs."<sup>5</sup>

Here, we are more upstream by the fact that it concerns the studies and not the operations and the functional mission itself. We will see more that the technico-operational (or functional) approach that we will define is at the bottom of the pyramid.

Staff profiles :

1. The technico-operational experts<sup>6</sup> of a domain working for example in a Technical Department whose mission is:
  - a. to capture the operational need and translate it into Technico-Operational requirements,

---

<sup>5</sup> <https://www.ixarm.com/fr/Presentation%2C88413>

<sup>6</sup> <https://fr.linkedin.com/jobs/view/expert-technico-operationnel-cyber-securit%C3%A9-h-f-at-thales-1621909449>

4

- b. to ensure the relevance of the technical content with respect to the technico-operational need
2. the specialists of the means of simulations<sup>7</sup>
- a. Ensure the implementation and the technico-operational follow-up of simulation means.

This category of staff will then be found in organizations such as the CATOD (Centre d'Analyses Technico-Opérationnelle de Défense = Center for Technico-Operational Analysis of Defense) which is the establishment that provides the organizations of the Ministry of the Armed Forces with decision-making assistance in the areas of forecasting, analysis, expression of need, realization and employment of weapon systems, in the form of studies or experiments."<sup>8</sup>

Nowadays, we must not forget the digital transformation within the organizations, and for that it is essential to "bring the IT department closer to the business" and "the new digital leaders have largely overcome the mutual distrust between business and IT." Indeed, "recruiting dual profiles" such as "technico-functional expert...will also be a good way to create bridges between departments with often incompatible training, languages and practices." [7]

The developer Benjamin Delespierre in the website [www.medium.com](http://www.medium.com) presents the cleavage that exists between technical and operational staff and, in certain cases, this can lead to the formation of clans within the company.

Clans whose members hate each other on principle: «« in any case, they understand nothing of what we do. It's not worth talking to them ». When you get to that point, there's no point in talking about collaboration and synergy within the company. I've seen it many times, there is little mixing of technical and non-technical people.

You don't have to be a management expert to understand that the project, and by extension the entire company, will suffer in the end. »<sup>9</sup>

But some people choose to combine the technical with the operational and/or functional in a submarine mode or in full visibility depending on the organization to which they belong.

---

7 REM defense : [https://www.defense.gouv.fr/re/content/download/538408/9248195/REM\\_2019\\_final.pdf](https://www.defense.gouv.fr/re/content/download/538408/9248195/REM_2019_final.pdf)

8 <https://www.defense.gouv.fr/dga/recrutement2/paris/paris-ingenierie>

9 <https://medium.com/france/les-d%C3%A9veloppeurs-sont-devenus-des-cols-bleus-457bd6287631>

Those people are considered as Shadows IT<sup>10</sup> and hackers because they do not respect the standards.

Nevertheless, they are useful.

With their corporate hacker<sup>11</sup> spirit, they show that there is often no ready-made tool to meet the need.

This includes both technical missions in an operational environment, known as "technico-operational", as well as the staff performing them.

Therefore, the second approach (the one we are interested in) is :

**Technico-operational (functional) approach:**

By staff we mean 2 categories:

- The first one being the staffs with a high level of expertise : specialists, technicians, engineers, operationals evolving in a technico-operational context and working very often in a business direction for which they have to carry out specialized missions requiring a particular technicality and expertise on the ground in operational situations (e.g.: analysis, data processing, ...).
- The second is the junior staff with a strong affinity for technology, information, languages, ..., who work in fast thinking and action mode and who then adopt what is called the system D.

Our observation of business needs allowed us to draw a first picture defining the technico-operational approach and a schema of reasoning and process of technico-operational (functional) approach (see Table 1).

*The following table gives a first definition of our approach :*

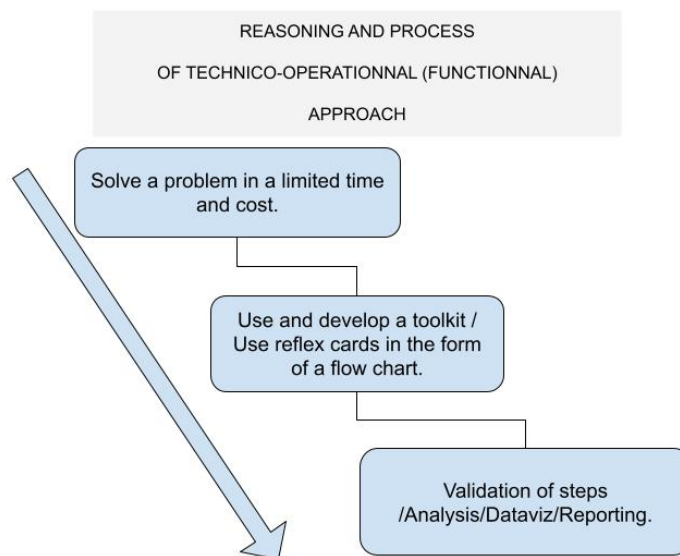
*Table 1 : Definition table for the technico-operational (functional) approach.*

<b>TECHNICO-OPERATIONAL (FUNCTIONAL) APPROACH IN NLP</b>	
<b>Description of the approach</b>	<b>Business applications of the approach</b>

<sup>10</sup> <https://www.lemagit.fr/definition/Shadow-IT>

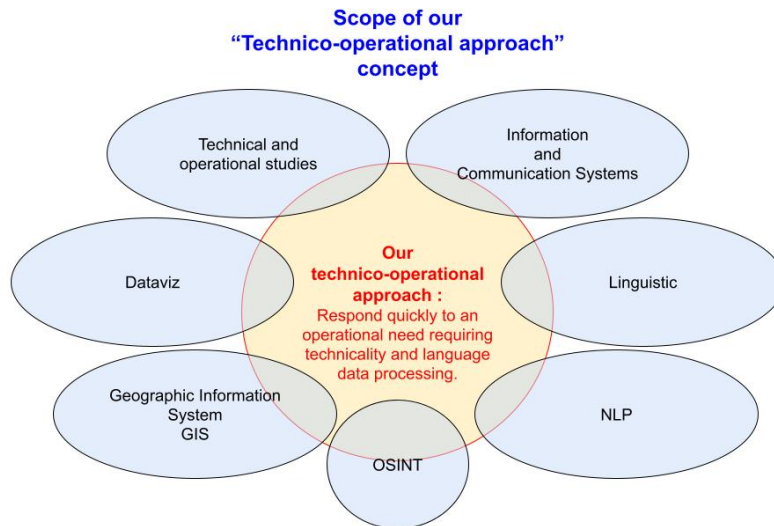
<sup>11</sup> *Understanding corporate culture hacking a new-found source of pleasure and efficiency in the workplace*, 2018, WAVESTONE.

- 
- Missions
  - People
  - Technical
  - Functional
  - NLP
  - Linguistic
  - Generally new themes
  - Context requires a strong reactivity and adaptability
  - Quick execution of tasks
  - One-shot task
  - Limited time
  - Limited cost
- Prospecting
  - Investigations
  - Geopolitical analysis
  - Discourse analysis
  - Text analysis
  - Analysis of poorly endowed languages
  - Knowledge Management
  - OSINT
- 



**Fig. 1.** Reasoning and process of technico-operational (functionnal) approach.

Regarding the technico-operational approach, it is important to accept to break silos even if some say that IT shodows can create data silos, in order to achieve a certain technical + operational/functional transversality without harming the specialists' fields as described in Fig. 1.



*Figure 2.* Scope of our « technico-operational approach » concept.

### 3 Case-study : NOOJ in a technico-operational context

A recent experiment in press analysis on "Qualitative Study of Dalian Wanda-AMC" presented in Sisi Zhao's Master's thesis [25], provides feedback on the use of NOOJ in a technical-functional context (press analysis) very close to our work which states that : "Overall, this researcher's experience with NOOJ was mixed: good for suggesting paths of inquiry and possibly seeing things not seen before, but not necessarily useful when doing a qualitative analysis, which still requires deep thinking about the text". [25]

We agree that NOOJ is a good tool for suggesting lines of inquiry and possibly seeing things not seen before. This is one of the reasons why we chose NOOJ technology.

However, regarding the point of Sisi Zhao that NOOJ is not necessarily useful to do a qualitative analysis, which always requires a deep reflection on the text. This is normal because NOOJ is a tool to formalize the language. In any case NOOJ requires an entry cost in the analysis of the language corpus on which we are working.

With the objective of testing NOOJ in a technical/operational/functional context, we decided to work on the Vietnamese language which is a poorly endowed language:



- very few annotated corpora.
- Vietnamese is a tonal and monosyllabic language.
- no verbal tenses.
- tense, aspect and modality markers.
- Compatible with Temporal Frames according to Michel Charolles' definition [4].

Our corpora will be from the Vietnamese online press.

### 3.1. Spatio-temporal treatment in a technical-operational/functional approach : the case of the Vietnamese press

**The Vietnamese language has :**

Tenses :	Aspect
Past	Accompli (đã)
Present	Progression (đang)
Futur	Iterative (V+lại)
Neutral	Terminative (xong, hết, thôi, etc.)
	Imminente (sắp)

Some illustrations :

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 washing TV, the phone rang.)

Tense : Past and aspect : đang = action is “during”.

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

sẽ = action is after.

(From Cao Xuan Hao (1998) & Danh Thanh Do-Hurinville (2009)) [6][14].

In our process we will use the spatio-temporal Frames according and discourse frames and their boundaries from Michel Charolles [4][5].

His conceptualization can be easily implemented in the linguistic development environment NOOJ as the following syntactic grammar of aspectuo-temporal frames and markers (see **Fig. 3.**) :

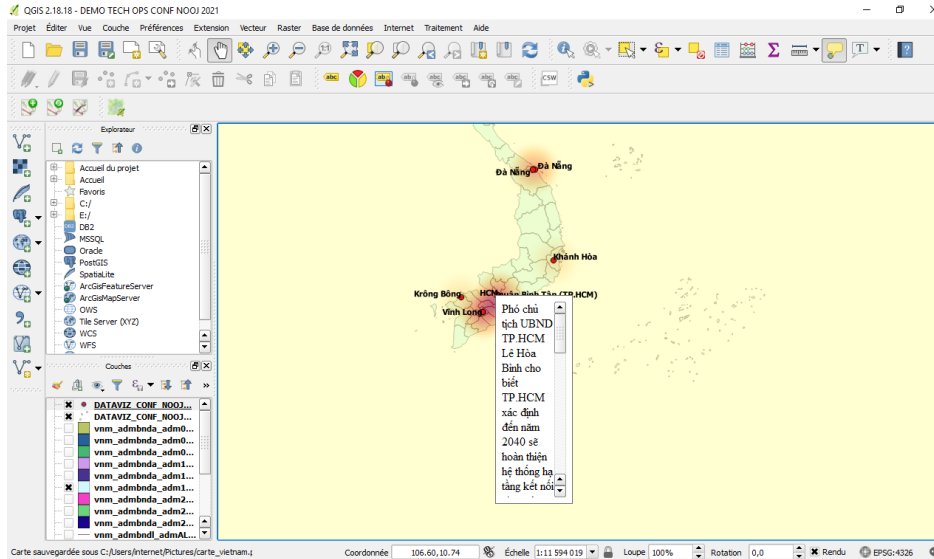


TEXT_UNIT	LOCALITEE_LOCALISATION	COUNTRY	Y	X	EVENT	field_7
1	Tam Giang	Vietnam	20.936853	106.339229	Kỷ luật chủ tịch xã mượn bằng n...	
2	Đà Nẵng	Vietnam	16.06778	108.22083	Ông Hồ Kỳ Minh, phó chủ tịch UB...	
3	Vĩnh Long	Vietnam	10.087867	106.002811	Công dân nhập cảnh làm thủ tục ...	
4	Bình An	Vietnam	21.732568	105.396672	Ông Võ Thanh Quang, Phó chủ t...	
5	Đà Nẵng	Vietnam	16.060667	108.107302	Để tạo điều kiện thuận lợi cho tro...	Kéo dài thời hạn ...
6	Sùng Nhơn	Vietnam	21.732568	105.396672	Ngày 29.12, trao đổi với PV Báo ...	
7	Cư Dã	Vietnam	21.732568	105.396672	Liên quan những sai phạm trong ...	
8	Krông Bông	Vietnam	10.712699	105.118808	Cùng ngày, ông Lê Văn Long, Ch...	
9	HCM	Vietnam	10.75	106.66667	Phó chủ tịch UBND TP.HCM Lê Hồ...	
10	quận Bình Tân (T...	Vietnam	10.816516	106.677868	Chiều 27-12, trao đổi với Tuổi Trẻ...	
11	Vĩnh Long	Vietnam	10.087867	106.002811	Ngày 27-12, Đoàn công tác của ...	
12	Đà Nẵng	Vietnam	16.060667	108.107302	Để tạo điều kiện thuận lợi cho tro...	Kéo dài thời hạn ...
13	Vĩnh Long	Vietnam	10.087867	106.002811	Ngày 26.12, ông Lữ Quang Ngườ...	
14	Vĩnh Long	Vietnam	10.087867	106.002811	Ngày 26.12, ông Lữ Quang Ngườ...	
15	HCM	Vietnam	10.75	106.66667	Chủ tịch UBND TP.HCM Nguyễn T...	
16	HCM	Vietnam	10.75	106.66667	Phát biểu tại buổi lễ công bố ngh...	

**Fig. 5.** - Table of events geolocated announced by the chairmen and vice-chairmen of the people's committees of different localities in Vietnam.

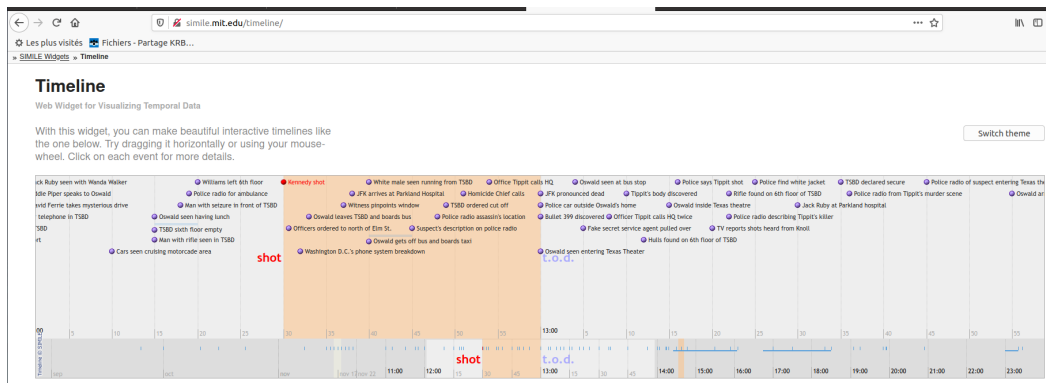
And then, we process this information in the free software QGIS<sup>12</sup> in order to have a mapping of the events as in following **Fig. 6**.

<sup>12</sup> <https://www.qgis.org/en/site/>



**Fig. 6.** – Map of events of the chairmen and vice-chairmen of the people's committees in Vietnam by using QGIS.

For your information, we used to do the chronological visualization with the open source solution SIMILE TimeLine (see **Fig. 7.**) which accepted XML and JSON files. Unfortunately, although it still works, the library is no longer maintained.



**Fig. 7.** <https://www.simile-widgets.org/timeline/>

Finally, this use case demonstrates the operability of NOOJ in a context of operational analysis of the Vietnamese press combined with geographic information processing.

### 3.2. Knowledge Management System (KMS)

#### **Knowledge Management and Personal Knowledge Management : definitions and applications**

##### *Definitions*

When we refer to Knowledge Management, the first problem we face is the definition of the concept. Since the 1960s, management sciences have been trying to define theoretical frameworks for valuing the processes of moving from data to knowledge. Let's give some of them here in order to show what they can overlap:

For Duzert, "The company's knowledge in all areas of expertise (research and development, production, purchasing, marketing, accounting, finance, etc.) is becoming a resource to be managed in the same way as a fleet of machines or the skills of employees. In order to face this challenge, companies have to set up specific management approaches called Knowledge Management (KM).

One of the most authoritative definitions comes from Jean-Yves Prax who breaks down Knowledge Management into four functional definitions:

1. Utilitarian definition: It defines Knowledge Management as a solution to the overabundance of information characteristic of the Web. It allows to provide a personalized and relevant answer in due time, without the request having been clearly explained (answer to a tacit request).
2. Operational definition: It is about combining knowledge and know-how in processes, products, and organizations to create value. In other words, the improvement of a company's performance requires the combination of knowledge (explicit and diffusible) and know-how (tacit and non diffusible).
3. Functional definition: This is the most widespread definition, which consists of describing Knowledge Management through the knowledge life cycle (formalization, validation, dissemination, etc.) and the devices necessary for its implementation (collaborative work, project management, EDM, etc.). According to Jean-Yves PRAX<sup>13</sup>, this definition is limited because it specifies how to do Knowledge Management and not what it is and why to do it.
4. Economic definition: It is a question of developing the intellectual capital of the company, i.e. all the knowledge and know-how, and of exploiting the skills already present, in order to improve its productivity.

---

13 Prax, *Manuel du knowledge management*.

A definition that would correspond more to our approach comes from Jean Michel<sup>14</sup> who distinguishes four distinct processes feeding the KM logic:

1. a process of collecting data, information, knowledge rules, feedback and other reports,
2. a process of modeling and extracting meaning, reformulation, elaboration of synthesis, creation of knowledge trees; it is a matter of moving from I (Information) to C (Knowledge) and
3. a process of collection, capitalization, and conservation of information
4. a process of redistribution and sharing with responses in either "push" or "pull" mode
5. **finally, a process of dynamization-regulation of all the previous processes with a concern for evaluation".**

#### *Personal Knowledge Management*

With Personal Knowledge Management, we move to another scale, one where the human being is truly put back at the heart of the knowledge equation.

Like Knowledge Management, there are a plethora of definitions, the two main characteristics of which are: 1/ a conceptual framework that organizes and integrates the information that we consider important and that should be part of our knowledge base and 2/ a strategy for transforming what can be considered as pieces of information that extend our personal knowledge base.

David Pauleen<sup>15</sup> from the University of New Zealand explains very well in an article dedicated to this theme that the basis of PKM is the notion of knowledge worker developed in 1998 by Davenport and Prusak<sup>16</sup>. On this basis, Jefferson will explain in 2006 that the technological revolution, with its uninterrupted flow of information, has made Personal information management systems (PIM) pivot towards PKM systems. First in a learning logic, since it is at the university that these issues became more pressing, then, more globally, in a managerial logic affecting all fields.

One of the important theoretical bases of KM is given by Nonaka and Takeuchi<sup>17</sup>, with two key dimensions: tacit or subjective knowledge and explicit or objective knowledge. In the technical-operational approach we present, our effort has been

---

14 Michel, « Le knowledge management, entre effet de mode et (ré)invention de la roue... »

15 Pauleen, « Personal Knowledge Management ».

16 Davenport et Prusak, *Information ecology: Mastering the information and knowledge environment*.

17 Nonaka et Takeuchi, *The knowledge-creating company: How japanese companies create the dynamics of innovation*.

focused on transforming tacit knowledge into actionable explicit knowledge for intelligence analysis.

The PKM system we used is based on two software bricks: a first brick for information capitalization and another brick for automatic language processing.

Since 2018, several PKM software have emerged with specific information processing functionalities, the main ones being search, categorization, structuring and integration. We can mention solutions such as Roam, Obsidian Bear or Notion that have met their public among the dozens of others available on the market.

### 3.3. Applying NooJ Linguistics Analysis throughout PKM frameworks

#### The framework

**The main logic behind the implementation of this case study is the possibility for analysts to have a tool for capitalizing on information for their operations.** The implementation of this tool had to meet three main constraints that we set ourselves in order to reproduce as accurately as possible the professional environment of the analyst:

1. An easy handling of the system via a turnkey system, intuitive and can be set up quickly.
2. Minimizing the risk of being hacked by offering the analyst the possibility to host the solution on his own server with controlled data flows. SAAS type software was therefore not privileged for our case study.
3. A system with functionalities that facilitate the exploitation of information, in this case through the use of a pivotal format for processing data (markdown) and specific software functionalities (visual representation of information for a Social Network Analysis)

#### Data processing for exploitation purposes

##### *The corpus*

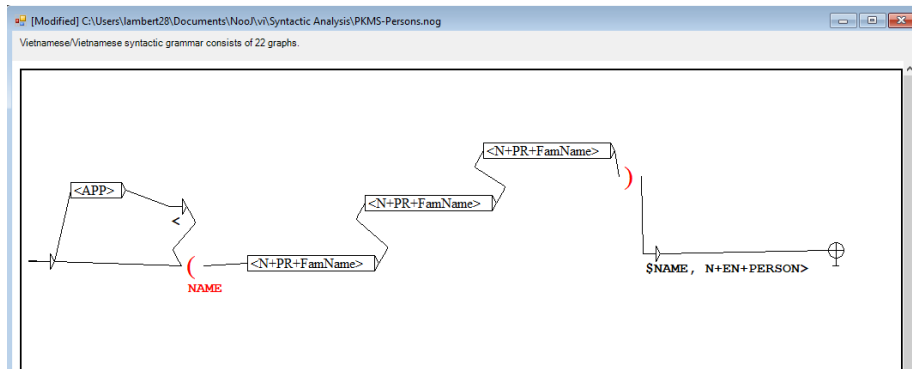
The test is based on a continuous exploitation of articles from the Vietnamese general press. The RSS feeds of three major newspapers of the National Press (Tuoi Tre, Thanh Nien, Nguoi Lao dong) were browsed daily over a week in order to extract the most recent articles. Each news feed had between ten and forty articles for a total of +1.5K documents.

The corpus obtained is a generalist corpus, including the information that the Vietnamese propaganda authorities wanted to highlight throughout the week. One of the objectives of the processing described below is to be able to quickly detect the major current affairs themes of the Vietnamese society.

#### *Preparation of the corpus and processing of the corpus*

The corpus used is composed of +1.5K raw text coming from the extraction of the RSS feeds of the three above-mentioned sites to extract three variables: the date, the author, the content of the article. These variables are reintegrated into XML files, which allows NooJ to process specific nodes. In parallel, with the Pandas library, a DataFrame is created gathering the documentary source and its variables. It will then be completed with the results of NooJ extractions (see below).

For this case study, we have created two types of linguistic resources: a specific dictionary with more than 15K entries, each of which is made up of the lexeme, its grammatical nature and a series of labels linking the term to its theme. For example, for proper nouns, we have two resources: the first is a dictionary with the first and last names of political figures as entries, followed by their functions and then the state organization to which they belong. The second resource is a grammar that extracts proper names (see Fig. 8.) The linguistic resources we have developed allow us to extract three types of named entities: **proper nouns, dates, and places** (see Fig. 9).



**Fig. 8.** Extraction graph for Proper Nouns



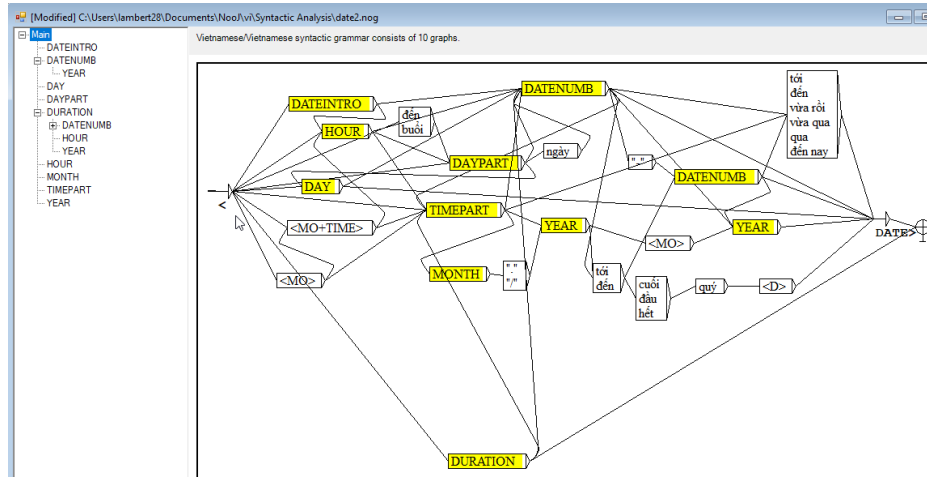


Fig. 9. : Graph for NER related to places

### The Markdown format and the Obsidian software solution

Markdown is a lightweight markup language created in 2004 by John Gruber with the help of Aaron Swartz. It offers an easy to read and write syntax. A document marked up with Markdown\*\* can be read as is without giving the impression of having been marked up or formatted by special instructions.

Concerning the information capitalization and visualization software layer, we chose Obsidian which is a new cross-platform PKM software allowing a complete management and visualization of relational links between several notes. The main advantage of this fast growing software is that it has a very active community of developers and that the texts produced can be placed on a local machine or a synchronized server, which considerably reduces the risk of data piracy.

The results of the NooJ extractions are added to the python dataframe which then allows the production of markdown files integrating these results. A markdown file consists of a YAML header (data serialization format) including the file creation date and the keywords (see Fig. 10).



```

---
tag : [Báo_Người_Lao_Động, Bộ_Công_an, Ban_chỉ_đạo, Covid-19]
url : [https://nld.com.vn/thoi-su/tam-thoi-go-bo-phong-toa-block-c-
chung-cu-bo-cong-an-o-tp-thu-duc-20210609211331865.htm]
---
Tạm thời gỡ bỏ phong tỏa Block C chung cư Bộ Công an ở TP Thủ Đức
Tối 9-6, trao đổi với phóng viên Báo Người Lao Động, lãnh đạo UBND phường An Khánh,
TP Thủ Đức cho biết đã gỡ phong tỏa Block C chung cư Bộ Công an ở phường An Khánh
theo hướng dẫn từ Trung tâm Kiểm soát bệnh tật TP HCM (HCDC)

"Ca mắc Covid-19 này được xác định nhiễm trong thời điểm thực hiện cách ly tập trung,
không phải tại chung cư. Những người tiếp xúc gần với bệnh nhân cũng được đưa đi
cách ly theo quy định. Theo hướng dẫn của HCDC, chúng tôi đã gỡ phong tỏa tạm thời
một phần bị phong tỏa tại chung cư này" - lãnh đạo UBND phường An Khánh thông tin.

```

**Fig. 10.** : Example of a typical file consisting of a YAML header including the tags extracted from NooJ

Using NooJ with Obsidian allows the analyst to structure the information he collects, to link the notes with the backlinks from the wiki logic and to produce knowledge based on the visualization of the information. There are specific Obsidian plugins that allow the processing of the named entities we extracted with NooJ: a timeline production plugin and a geolocation plugin that allows to integrate geographical maps from the GPS points of the named entities of places. Linked to the textual graph, the system offers a dynamic representation of the knowledge produced.

The other interesting point of the system is that the analyst is at the heart of the process of structuring the information he collects to produce intelligence.

## 4 Conclusion

To conclude, we were able to lay the first stones of the definition and conceptualization of the technical-operational/functional approach in NLP. This allows us to open the bibliographic field on this approach.

Following our experimentation, we note that NOOJ is compatible with a technical/operational/functional use in the context of discourse analysis and knowledge management.

However, NOOJ requires some expertise, for example when implementing models like Michel Charolles's discourse frameworks.

Nevertheless, a normal user can extract named entities without being an expert. But we would recommend him to follow the NOOJ course.

We can deal with poorly endowed languages vs AI and statistical model.

We should note that it was not possible for us to evaluate our grammars due to the lack of annotated corpora.

In a context that requires reactivity and strong adaptation, should we stop a system that has not yet been calibrated and evaluated? Or should we evaluate it in Agile mode, iteration by iteration or even *a posteriori*, especially if it has already fulfilled its mission of finding relevant information for which the user would have had difficulty finding it without tools such as NOOJ?

## 5 Perspectives

Since mid 2021, on the GITLAB repository, a new open source version of NOOJ written in SWIFT <https://gitlab.com/Silberz/RA> is being created.

Indeed, the university professor Max SILBERZTEIN has taken the option to put a multi-OS compatible version to the NLP community.

We are convinced that this new version of NOOJ named RA will allow to integrate into a real TOOLKIT that will help people working in an operational/functional context.

To overcome the level of expertise that would be required to use such a

However, RA will need to interface with several open source technologies in order to be able to do fully automated end-to-end processing, and then provide some modularity to RA.

To overcome the level of expertise that would be required to use such a toolkit, it will be necessary to foresee reflex cards based on logigrams but also to conceive and conceptualize the construction of NOOJ grammars by using, for example, blocks like in Pythons for Kids<sup>18</sup> or Scratch<sup>19</sup>.

Finally, we could use the Analysis of Competing Hypotheses (ACH)<sup>20</sup> whose challenge would be to see if we can filter the ACH results with NooJ grammars to corroborate the hypotheses.

---

18 <http://jasonrbriggs.com/python-for-kids/>

19 <https://scratch.mit.edu/>

20 ACH : « Analysis of competing hypotheses, sometimes abbreviated ACH, is a tool to aid judgment on important issues requiring careful weighing of alternative explanations or conclusions. It helps an analyst overcome, or at least minimize, some of the cognitive limitations that make prescient intelligence analysis so difficult to achieve. », Chapter 8 Analysis of Competing Hypotheses, Psychology of Intelligence Analysis by Richards J. Heueur, Jr. Center for the Study of Intelligence, CIA, 1999.

## 1 References

1. Bao Quoc Ho, Chevallet J. , et Bruandet M. , « Mise en place d'un Système de Recherche d'Informations en Vietnamien », in TALN 2003, 2003, 12.
2. Boffo N. et Ho Dinh O., « Automatic Processing of Temporality for VIET4Nooj », in NooJ 2010 Conference in Komotini (GR), 2010.
3. Boffo N., "Approche technico-fonctionnelle avec NOOJ : Le cas du traitement spatio-temporel de la presse vietnamienne", semaine NOOJ 2021 à l'INALCO.
4. Charolles M. , "L'encadrement du discours : Univers, champs, domaines et espaces" , 1997, HAL Id : hal-00665849, version 1.
5. Charolles M. , "Les cadres de discours et leurs frontières", 2012, HAL Id : hal-00665820, version 1.
6. Do-Hurinville D. T., in "Temps, Aspect et Modalité en vietnamien" (2009).
7. Ducrey V., Vivier E., Le guide de la transformation digitale, La méthode en 6 chantiers pour réussir votre transformation , 2019, Collection Hub management, Eyrolles.
8. Foucard J-L., La boîte à outils du Pilote des Systèmes d'information, 2ème Edition, DUNOD, 2014.
9. Kim Phuong " Ngu phap tieng viet nhung van de thoi, the "(2008).
10. Lambert P. et Boffo N. , Modules NOOJ de traitement de la langue vietnamienne, <http://www.nooj-association.org/resources/vi.zip> (dernière mise à jour janvier 2021).
11. Lambert P. , Fournié M. , et Ho Dinh O., « VIET4Nooj A Vietnamese module for Nooj », in NooJ conference 2010, 2010.
12. Lambert P. , Schwer S. , Boffo N. , “A new model of Time Expressions Detection and Annotation in Vietnamese : The hòm case” in International Conference on Asian Language Processing, 2012, Hanoi.
13. Lambert P. et Sidhom S. , « Using NooJ in a Process of Economic Intelligence: Producing Knowledge for Information Monitoring and Re-indexing Content », in NooJ 2010 Conference in Komotini (GR), 2010.
14. Linguistic approach of Cao Xuan Hao.
15. Nguyen T. B. et al., « Developing tools and building linguistic resources for vietnamese morpho-syntactic processing », 2004
16. QGIS, Système d'Information Géographique Libre et Open Source : <https://www.qgis.org/fr/site/index.html>
17. Schwer S. , Boffo N. , Lambert P. , "Toward Time Universals Identification in NooJ : A New Model of Time Recognition in Vietnamese", NooJ 2013 International Conference, Saarbrücken - June 2013.

18. Serrano L. , "Vers une capitalisation des connaissances orientée utilisateur : extraction et structuration automatiques de l'information issue de sources ouvertes", 2014, HAL Id : tel-01082975, version 1.
19. Silberztein M., "La formalisation des langues, L'approche de NooJ", ISTE 2015.
20. Silberztein M. , « Les dictionnaires DEM et LVF au service des humanités numériques », Linx [En ligne], 80 | 2020, mis en ligne le 10 juillet 2020, consulté le 05 août 2020. URL : <http://journals.openedition.org/linx/6072> ; DOI : <https://doi.org/10.4000/linx.6072>
21. Silberztein M. , « NooJ: A Cooperative, Object-Oriented Architecture for NLP », IN-TEX pour la Linguistique et le traitement ..., 2004.
22. Silberztein M. , NOOJ V4, 2020, HAL Id: hal-02435923.
23. Zakaria Kurdi M. , "Traitement automatique des langues et linguistique informatique 1 , Parole, morphologie et syntaxe", ISTE 2017.
24. Zakaria Kurdi M. , "Traitement automatique des langues et linguistique informatique 2 , Sémantique, discours et applications", ISTE 2018.
25. Zhao S., Cultural Exchange and Media Evaluations Behind Transnational Business Acquisition Between China and the United States: A Qualitative Study of Dalian Wanda-AMC ,01 October 2018.