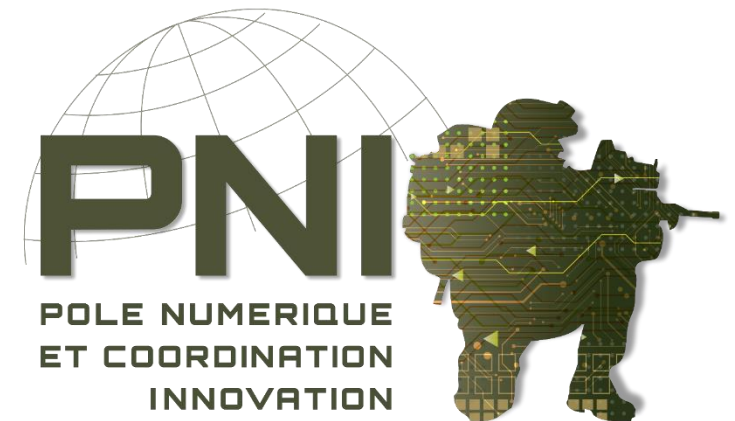




Improving understanding by combining NLP and Business Analysis techniques





Who are we ?

Analyse and operational research departement :

Lieutenant-Colonel Bass (Polytechnique)

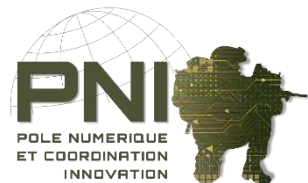
Capitaine Calligaro (Telecom Paris)

Main task :

Cartographic and visual representation

Build improved data for better understanding of the warfield

On demand, create a data visualization to enhance insights





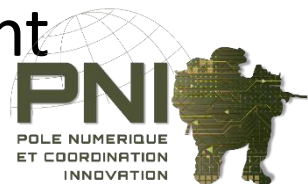
Purpose

Events are described by two kind of data, numerical and open text.

Dashboards are usually made using numerical data.
Quite often open text is basically analyzed by staff officer.

Therefore, leaders don't have a full understanding of the event and this process is costly

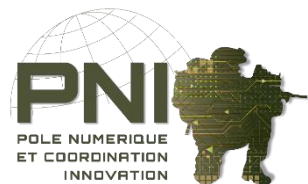
We want to improve the leader's understanding with a cost efficient report covering both data sources





Agenda

- Overview of 2 old challenges
- Deeply focus on a new challenge
 - Natural Language Processing approach
 - Recurrent Neural Network approach
 - Results and insights
- Conclusion





Challenge 1



Maintenance : improve down time (MDT) prediction
thanks to analysis of repair acts comments (GEN TXT)

Purpose : to predict the « waiting for parts » time (which can be a significant part of total down time)

First idea : it depends on parts references involved across all repair operations during the same « down » phase

But ... the first model has to be (dramatically) improved ...

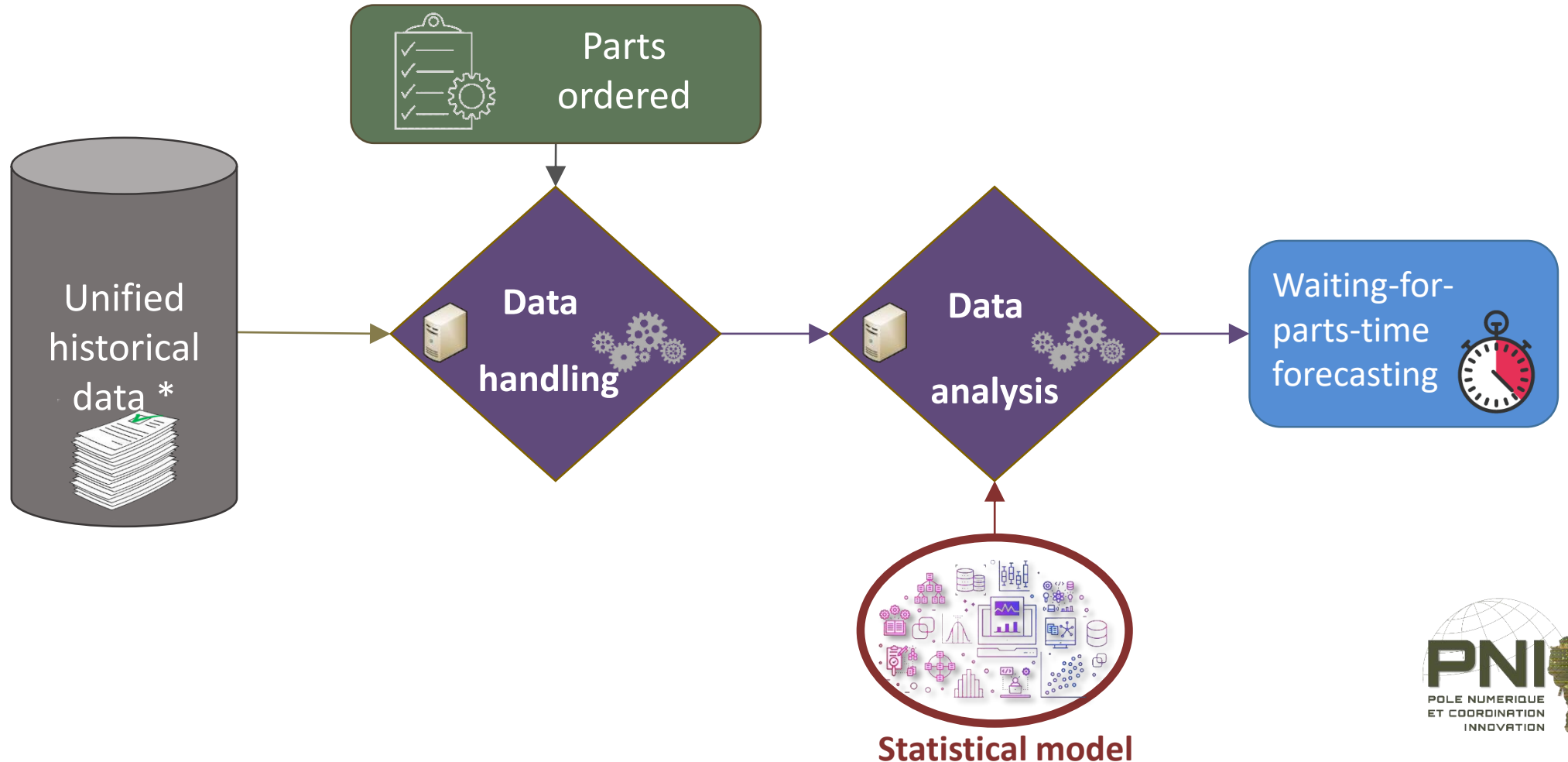
... thanks to NLP !





General purpose : modelizing the maintenance of major combat vehicles

- ✓ Use case : modelizing the relation between waiting-for-parts time and which parts have been ordered during one same « down period » (for all-day maintenance)





Result : a rather poor model ...

- ✘ A basic machine learning is not enough :
 - Too many different references (2400 items over 5 yrs)
 - An ill-balanced data set (rare major events are underweighted but significant for higher down times : major parts !!)
- ✘ Still significant forecasting errors :
 - +/- 11 days compared with a time ~ 46 days for this data set
 - Which parts have been ordered can not totally explain the waiting time ...

...without integrating GEN TXT exploitation, which is slightly more difficult ...

- ✘ Various data formats :
 - ↳ 19-1214RMAT7LON / 1SPAHIS / VBL n°69240035 / VSEB
- ✘ Various levels of details in repair act comments :
 - ↳ REPARATIONS ++ Collecteur échappement HS + Poulie refroidissement alternateur HS + courroie alternateur manquante + joint spy reducteur AVG HS + fuite vanne thermo + réglage frein de parc + protéger flexible frein AV + remettre vis sur protection phare IR + graisser palier direction SUP G
- ✘ Very erratic data quality :
 - ↳ VS 1 an + ctrl FT + COS

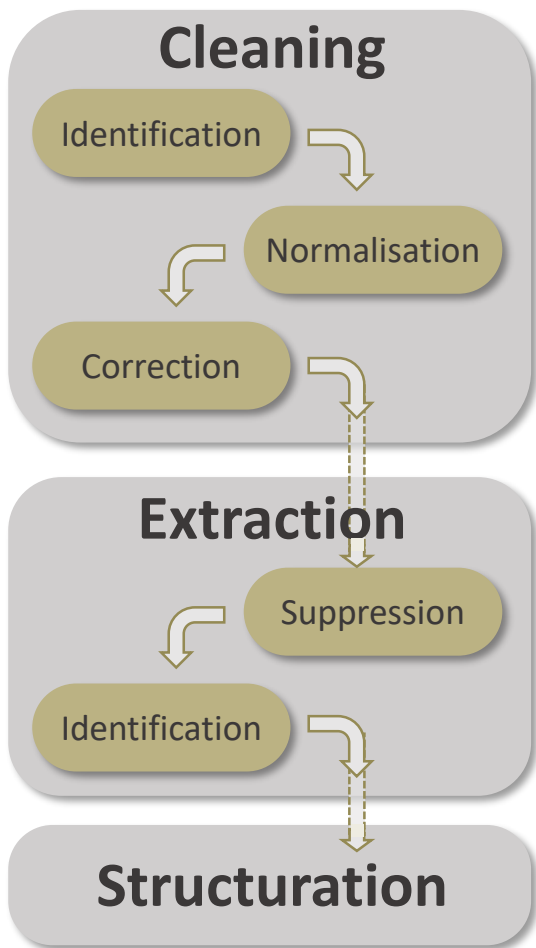
One man, one dialect !





Grabbing value from GEN TXT with AI

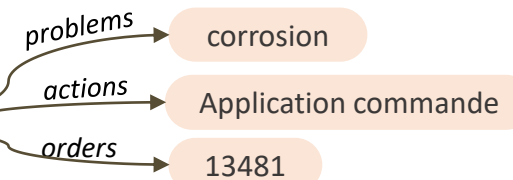
A method has been created to perform following steps :



- * Identify the categories related to FR MoD
↳ *Units, equipments, events, operations ...*
- * Normalize maintenance expressions
↳ *acronyms, problems, actions, parts ...*
- * Get a corrected and unified semantics
↳ *Correct errors and extract lexical domains (through AI)*

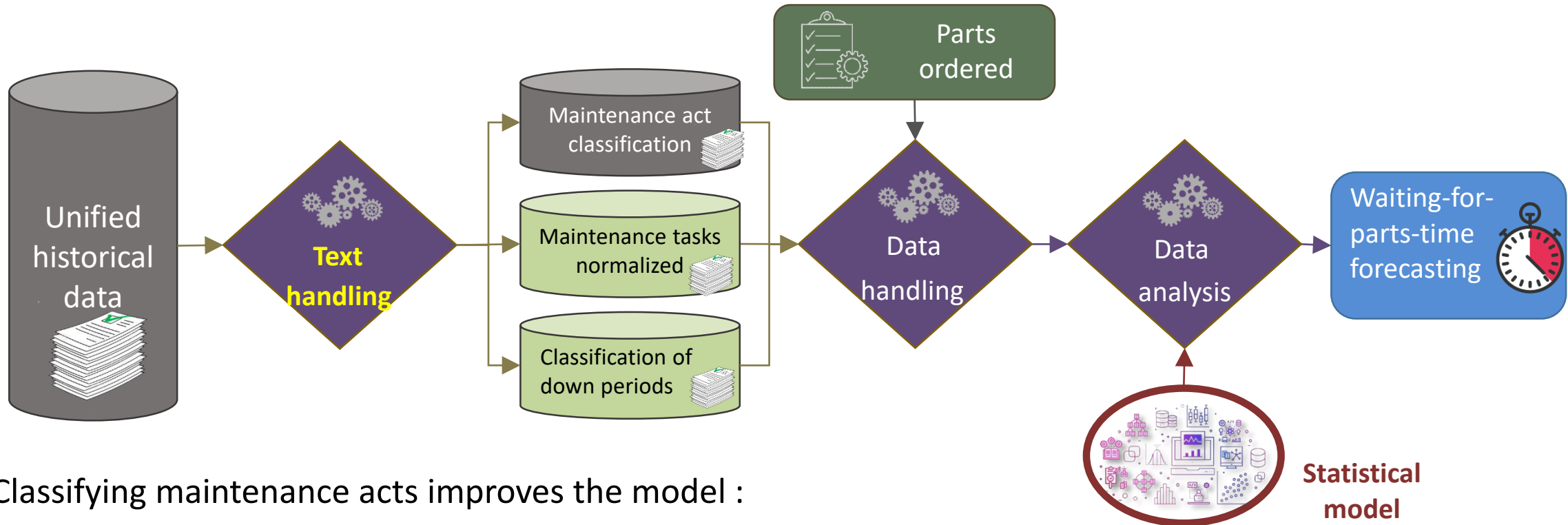
- * Erase informations already given by structured data
↳ *Repair act # / Vehicle # / Type of equipment / Repairing unit / ...*
- * Identify data of interest
↳ *Maintenance operations frequency, conducted actions, technical orders, normalized maintenance tasks, ...*

application de la FT 13481 ++
Commande prévisionnelle du KIT
Extincteur avec début de corrosion



The information can now be exploited in models !





Classifying maintenance acts improves the model :

- + Forecasting error is lower (+/- 9j)
- + **The structured information is retained as #1 criterion !**

⇒ **Way ahead :**

- ✓ Classify the type of « down » period (first cause ?)
- ✓ Maintenance tasks classification ?



Challenge 2

Improve psychological analysis

Pre-processing of short answers to an open-ended question

Who was
your main
support ?

Apprendre de nouvelles choses.
Apprendre à connaître différentes personnes, unités et spécialités.
Servir à quelque chose.
Le succès de la réalisation de la mission et le retour en bonne santé et sans blessures physique et mentale.

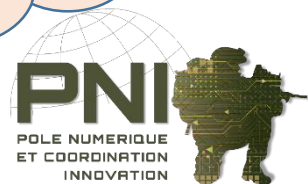
Ce qui m'a apporté le plus de satisfaction durant mon mandat était le fait qu'on me considère comme une personne à qui on pouvait demander un avis technique ou me donner des missions demandant une certaine réflexion.

Autonomie,
responsabilité,
liens amicaux.

Accomplir
la mission

Apprendre
aux jeunes

J'ai appris l'italien
et me suis
amélioré au Poker





Challenge 2 : context

A psychologist, provided with a form, must produce a study on the general psychological state of the military returned from mandate.

In this form, there will be an open-ended question and people will answer it with a short sentence.

We want to analyze and summarize all the answers, giving a first generalist approach to the psychologist.

The result will be a Data Visualization...



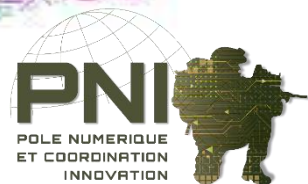


Word Cloud

A Word Cloud is a data visualization that turns text into an image.

The font is directly related to the importance of the word.

Shape, color and orientation are just a style choice.





Conclusion

Gain :

The specialist estimated that he saved 30% of the time on the analysis of the questionnaire

User feedback :

The specialist quickly gained confidence in the result.

He easily appreciates and absorbs the proposed "data viz"

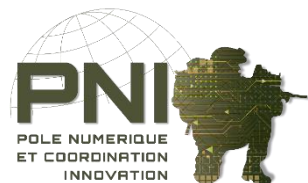
He considers that it gives him a good vision of insights

Conclusion :

The specialist must be accompanied in his work and not replaced.

Start with simple things then accumulate, do not try to do better than the specialist

The choice of data viz is essential

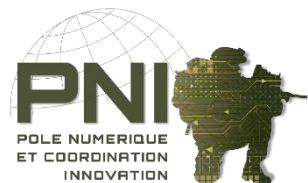




Challenge 3

Exploit data from old mission report
Produce structured data from open text

Report Date (Local)	Title	Description	Location Details	Actions Taken	Personnel Details	Equipment Details	KMs
2018/03/13 13:11:00	DETECTION INDIRECT FIRE BY RADAR COBRA	[21161607A] 9LAB COND CHEF 9 LAB REPORT who RICM Where: 33VWC5568092406 MANSAMALA When : 21MAR18 AT 1500Z 1. 1 VEHICULE VBL DESTROY 2. 10 line EOD report send by MAIL . 3. NO WIA NO KIA	LOCALISATION ENY 33 VWC 56614 / 82865 33 VWC 51916 / 76407 LOCALISATION IMPACTS 33VWC 27952 / 83111 33VWC 27611 / 84661	Recce	Ltn Duchateau Adj Lebonvieu Sgt Gouzman Sgt Anderson Sgt Chatillo	VBL 12 Famas VBL + Cobra Radar	200





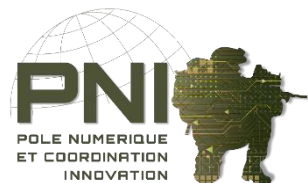
Natural Language Processing approach

The Natural Language Processing is a set of statistical method applied to text.

Used to extract information from text.

Two ways to analyze :

- The syntax of word (key word)
- The semantics of sentences (Part of Speech)





Key Word / Theme

- Group word by theme (ex : weapon, location ...)
- A specialist is mandatory to create lists of words and themes

070806ZMAR2018

08h04Z liaicellinsbctops From SBCT:

The company had an .50mm accidental discharge while attempting to clear a weapons jam.

The AD damaged a locals property.

The round went through the side of the house causing shrapnel that wounded a 12 year old girl.

The BDE CDR personally visited

the family to smooth things over.

The neighborhood in which the AD happen have been instrumental to HUMINT as it pertains to cell movements.

MEDEVAC requested and 9 line submitted.

Green : Location

Red : Weapon

Bleu : Civil

Purple : Relation

Orange : Military acronym



Part Of Speech

PoS : Grammatical identification of words within a whole sentence

Grammatical groups :

Noun, Pronoun

Verb, Adverb

Adjective, Conjunction

Preposition, Interjection





Part Of Speech

Category Code	Title	Description	Location Details
CIVRTA	TRAFF	REPORT FROM 1 RIMA - 121330ZMAR18 -A VBL OF THE 1RST ESC HAS GOT A TRAFFIC INCIDENT AGAINST A CIVILIAN VEHICULE IN THE TOWN OF KOSTA. A CIVILIAN FEMALE IS A LITTLE INJURED ;THERE IS NO MILITARY INJURED TWO CHILDRENS ARE SAFE IN THE CAR. THE VBL HAS A BROKEN WHEEL. REPORT/121429AMAR2018	KOSTA

Human	Destruction	
yes	Yes	Civilan Female
	No	Military Two Children
No	Yes	VBL
	No	

A CIVILIAN FEMALE IS A LITTLE INJURED.

THERE IS NO MILITARY INJURED .

TWO CHILDRENS ARE SAFE IN THE CAR.

THE VBL HAS A BROKEN WHEEL.





Conclusion on NLP

- 40% of the information can be found and extracted
- Requires continuous update of key words & rules ...
(vocabulary and rules are linked to activities operation !)
- Difficulties on :
 - Writing quality (spelling, grammar)
 - Coreference (*Paul arrived, He brought his shoes (who is the 'he')*)



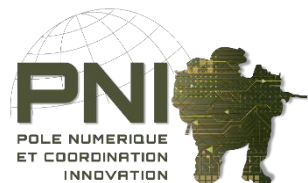


Recurrent Neural Network approach

Recurrent Neural network is the state-of-the-art algorithm for sequential data, where word order is important.

It is the first algorithm that remembers its input, thanks to an internal memory.

The one dedicated to text is named BERT.

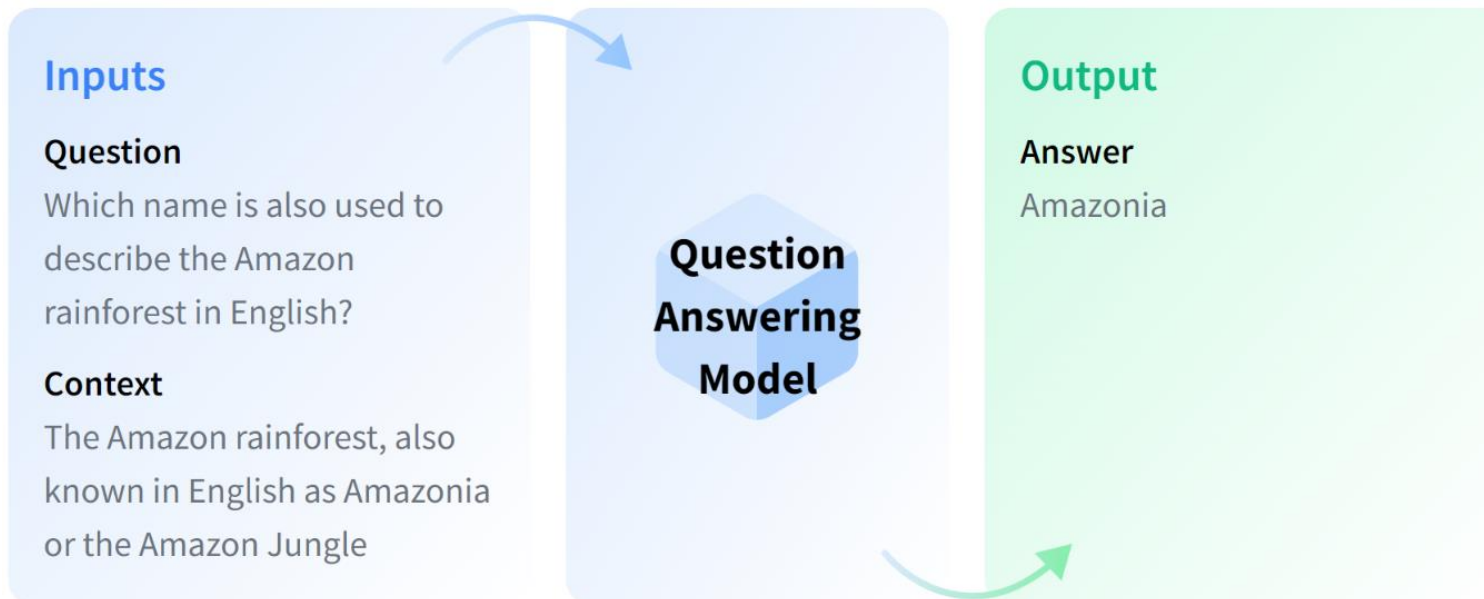


BERT and the Question Answering

- BERT (EN) / CamemBERT (FR)
- Question / Answering



CamemBERT





How we use Question Answering

We produce one question per structured data that we can / want to produce.

Some simple questions :

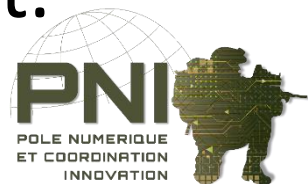
Where, When, What, who

Some more complex questions :

When did the accident / meet appen

Who was imply in the accident / meet

The difficulty is when a question is not relevant for a text.





An irrelevant question

- Text:

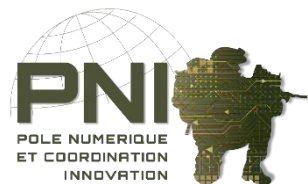
2 unknown males with binoculars observing own field radar. When approached by subunits, the car left in western direction. Probably red Ford focus. Licences plate unknown

- Question:

When did the accident happened ?

- Model's answer:

Binoculars observing own



Results on BERT & Question Answering

- 80% of accuracy on the best part of question set
Excluding the irrelevant question
- 20% of accuracy on the whole question set
include irrelevant question

The main difficulties encountered were :

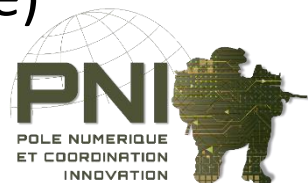
- Is the question is relevant ?
- Is the answer is coherent ?



Work on 'Is the answer coherent'

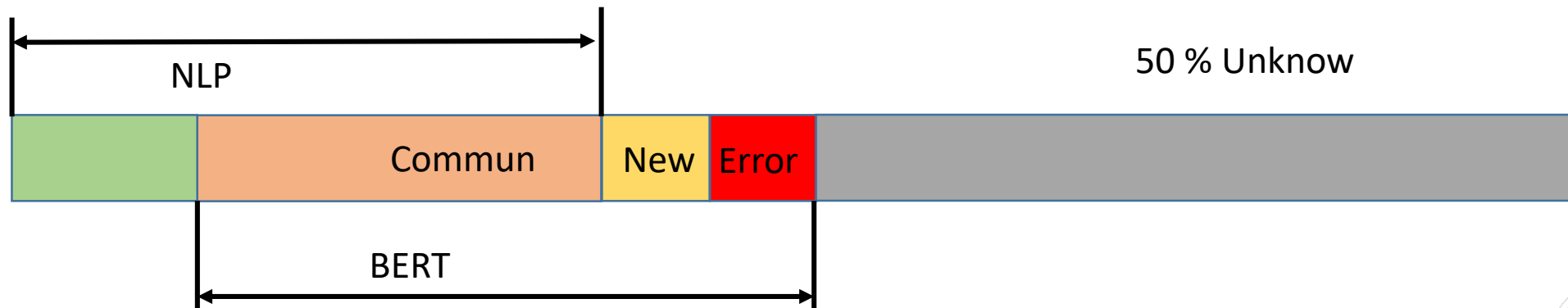
Theory and working basis

- Goal :
 - Identify and reject answers where the sentence is not syntactically correct
- How :
 - We will ask to a BERT model if it could have produced this sentence
- Prospect :
 - As for image recognition we will use a BERT GAN
 - With a BERT 'add word' we will check for the next word (End of sentence)



Conclusion on BERT and NLP

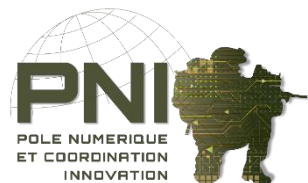
- 50% of information stays unidentified
- 10% provided only by NLP
- 30% identified by NLP & BERT
- BERT provides 5% more information and 5% more error than NLP





Way ahead

- Technological watch on Q&A models and their evolution
- Work on Automatic detection if an answer is consistent (GAN Bert)
- Automatic detection if a question is relevant





Final Conclusion

- All solutions are complementary
- NLP requires continuous update and control
- BERT goes further than NLP but the technology isn't yet fully matured
- Both processes are still costly

