

Comment exprime-t-on un problème technique ? Repérer, typer et formaliser

Séminaires "Thématiques actuelles de la recherche en TAL"

4 novembre 2024

Mariame Maarouf



Contexte et objectif industriel de la thèse



1. Recherche de solutions innovantes
2. Capitalisation des REX pour la résolution de problèmes techniques



1. Accompagne les entreprises dans la mise en place d'une méthode de capitalisation de connaissance → TRIZ
2. TRIZ : méthode de résolution de problèmes techniques

Objectifs de la thèse :

1. Traiter des problèmes techniques par le biais d'une modélisation "à la TRIZ"
2. Retrouver des problèmes similaires déjà rencontrés dans une base documentaire et ses solutions

Questions et problématiques de recherche

Problématiques :

1. Comment traiter un corpus technique et bruité avec des méthodes de TAL ?
2. Comment repérer des patrons récurrents d'expression d'un problème technique ?
3. Quelle représentation sémantique d'un dysfonctionnement technique ?
4. Comment identifier les informations pertinentes à l'application de TRIZ et comment les modéliser ?

Questions :

1. Comment exprime-t-on un problème technique dans un texte de type REX ?
2. Est-il possible d'émettre une grammaire d'un problème technique ?
3. Quels sont les éléments présents et comment sont-ils connectés ?
4. Est-ce que les REX sont des documents pertinents à l'utilisation de TRIZ ?

Départ et arrivée

1. Qu'est-ce qu'on a ?
2. Qu'est-ce qu'on veut ?

Qu'est-ce qu'un REX ?

REX (Retour d'EXpérience) : procédé de collecte d'information dans un but de maîtrise des risques

FA (Fiches Anomalies) : REX produits lors d'opérations de maintenance sur la fusée Ariane 5

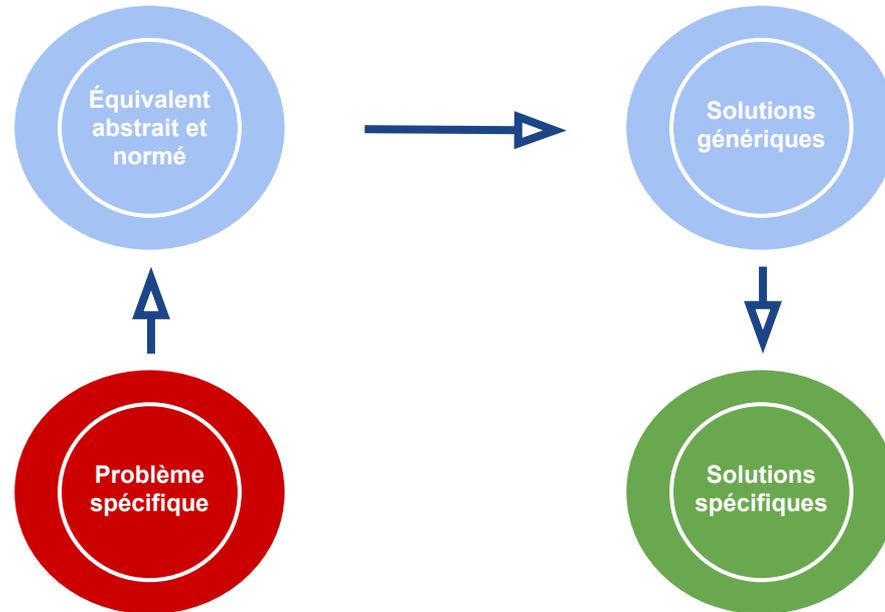
- Constat rédigé sur le vif
- Tout écart à l'attendu rencontré

Corpus :

- 35 242 FA produites entre 1997 et 2011
- Version minimale du problème

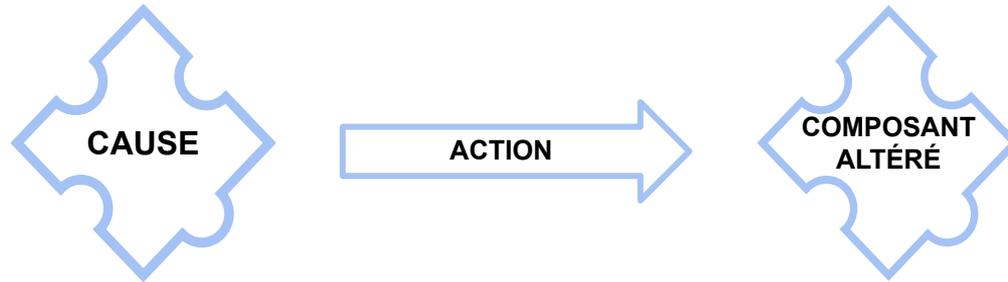
Qu'est-ce que TRIZ ?

- TRIZ (Théorie de Résolution des Problèmes Inventifs)
- Méthode de résolution de problème
- Développé par G. Altshuller en 1946 à partir d'analyse de brevets



Qu'est-ce que TRIZ ? - Modélisation

Formalisation du problème sous forme de *vépole* : modélisation d'une action physique entre deux éléments



Modélisation - point de départ

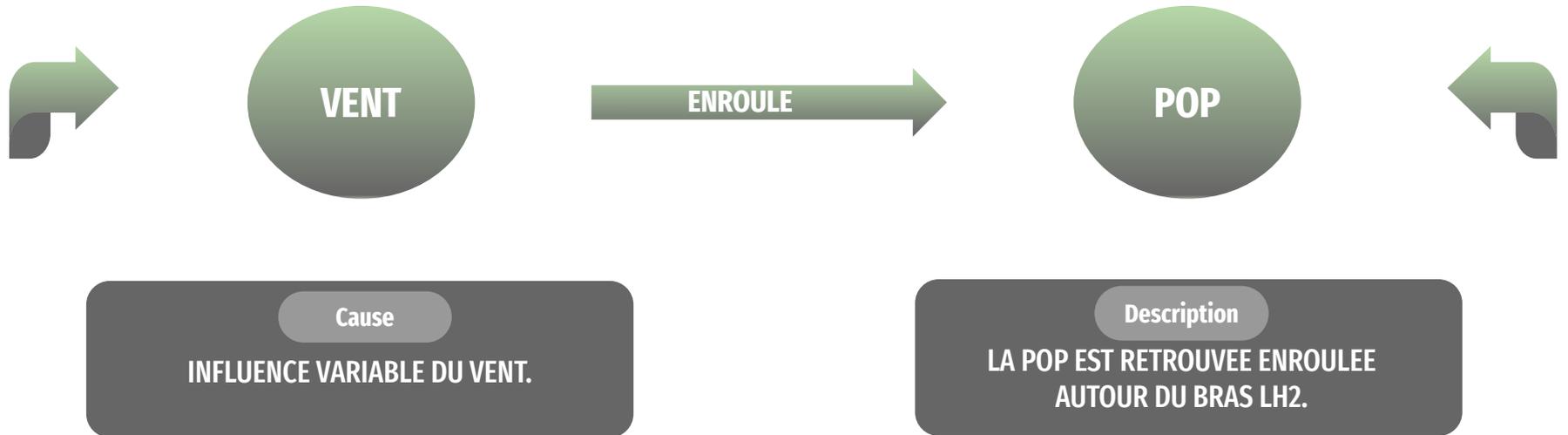
Cause

INFLUENCE VARIABLE DU VENT.

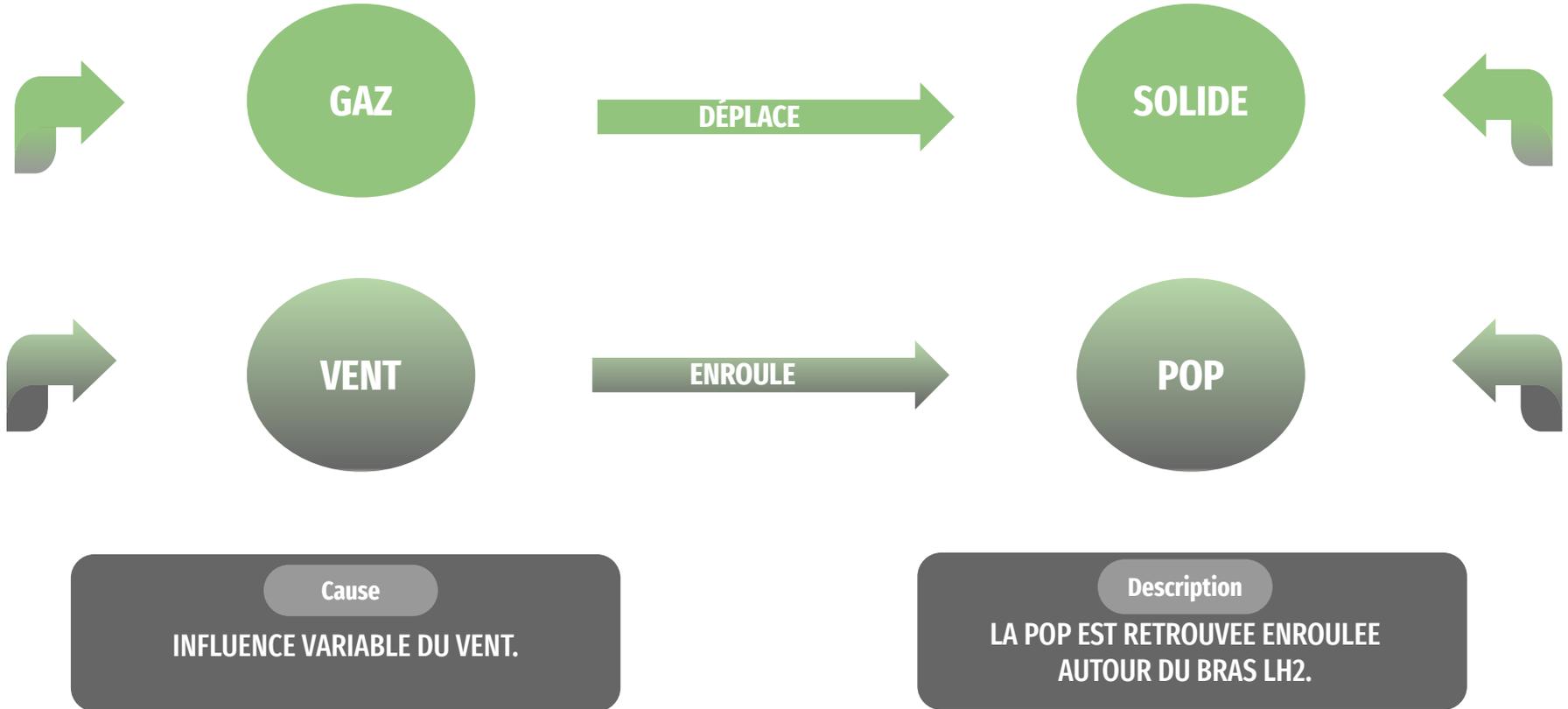
Description

LA POP EST RETROUVEE ENROULEE
AUTOUR DU BRAS LH2.

Modélisation - entre corpus et abstraction



Modélisation - point d'arrivée



Aperçu du corpus

Aperçu du corpus

Description du problème
DEFAUT BETON SUR CARNEAUX :1) DESSUS CARNEAU EST.JPG2) DESSUS CARNEAU NORD (1).JPG (AFFAISEMENT GENERAL)3) DESSUS CARNEAU EST.JPG
DEFAUT DE PLANEITE SUR LE CABLE X POINTS. ON CONSTATE UN LEGER SURPLUS (ENV. X,Xmm) SUR LA SURFACE du connecteur (CABLE EBM COTE OXYGENE : FFL3918452)
defaut vidange eclateur N°10 piscine SPI

Aperçu du corpus

Description du problème

DEFAUT BETON SUR CARNEAUX :1) DESSUS CARNEAU EST.JPG2) DESSUS CARNEAU NORD (1).JPG
(AFFAISEMENT GENERAL)3) DESSUS CARNEAU EST.JPG

DEFAUT DE PLANEITE SUR LE CABLE X POINTS. ON CONSTATE UN LEGER SURPLUS (ENV. X,Xmm) SUR LA
SURFACE du connecteur (CABLE EBM COTE OXYGENE : FFL3918452)

defaut vidange eclateur N°10 piscine SPI

Légende :

- Espaces et ponctuation
- Fautes
- Minuscules
- Accents
- Acronymes

→ Pré-traitement / nettoyage des données avec outils traditionnels
trop lourd

Que fait-on quand son corpus ressemble à ça ?

1. Est-ce qu'on observe des régularités dans l'expression d'un problème technique ?
2. Comment les repérer ?

Exploration du corpus

Termes les plus fréquents	
Terme	Fréquence
table	2561
niveau	2510
default	2346
eap	2174
hors	2128
perte	1998
ligne	1890
lh	1867
vis	1822
baf	1681

Termes relatifs aux problèmes les plus fréquents	
Terme	Fréquence
default	2346
perte	1998
fuite	1654
impossibilité	1146
absence	1118
impossible	1027
erreur	914
problème	487
fonctionne	466

Exploration du corpus

Freq	N-grams / taille = 2
926	impossibilité de
581	absence de
541	hors spec
536	perte de
516	au lieu
500	impossible de
499	présence d
482	fuite sur
469	perte du
447	présence de
438	hors specification

Freq	N-grams / taille = 3
435	au lieu de
372	ne sont pas
296	ne fonctionne pas
271	n'a pas
167	a l'arrêt
167	pour une spec
165	présence d'un
139	présence d'eau
113	d'eau dans
112	perte de l
106	impossibilité de monter

Exploration de corpus : Topic modeling

Rappel :

- Méthode d'apprentissage non-supervisé
- Clustering de mots
- Faire émerger des thématiques au sein de documents

Lien topic modeling :

file:///home/mariame/Documents/These/Données/topic_modeling/viz_11_topics.html

Typologie

1. Repérage de marqueurs lexicaux d'un problème technique
2. Élaboration d'une typologie d'expression d'un problème basée sur ces marqueurs lexicaux
3. Confrontation systématique entre la typologie et les données jusqu'à arriver à une version stable
 - a. Sélection d'échantillon(s) afin de vérifier la couverture de la typologie
 - b. Annotation manuelle d'un expert et d'un non-expert

Typologie

N°	Type	Exemples de marqueurs	Exemples de FA
1	Fuite	fuite, fuyard, écoulement	FUITE D'AIR A L'INTERFACE POPH/CASE.
2	Signal qui s'est déclenché	témoin, alerte, alarme	ALARMES DE DIALOGUE AVEC LES BALANCES PRECIA.
3	Obstacle	gêne, empêche, bloque	VANNE D'EVENT MAT VVC 9006 DE LA TABLE 2 BLOQUEE .
4	Dégradation - Usure - Saleté	cassé, marques, corrosion	LA 1ERE LANGUETTE DU SYSTEME DE DEPLOIEMENT 1/2 COIFFE TN EST CASSEE .
5	Élément manquant	absence, sans, manque	ABSENCE DES PEIGNES, PALETTE EAP2, DES FLEXIBLES CVPR, EPSR ET CVIRH.
6	Configuration hors spécification	hors spécification	FILETAGE DU CAPTEUR 22 CVM 9035T NON COMPATIBLE DU PAS DE FILETAGE
7	Dispositif qui ne fonctionne pas	HS, panne	NON FONCTIONNEMENT DE HVC199
8	Action difficile ou impossible	impossible, difficulté	IMPOSSIBILITE DE RACCORDER LE CABLE 21 WMT938 (SONDE DE TEMPERATURE)
9	État du monde		DECLENCHEMENT DE L'ASEPC.

Après, on fait quoi ?

1. Comment obtenir une représentation sémantique d'un dysfonctionnement technique ?
2. Comment repérer les éléments nécessaires à la modélisation en vépole ?

Frames (Fillmore, 1976)

Théorie sémantique des cadres (Fillmore, 1976)

- Signification d'un terme : contexte situationnel et événementiel
- Situation prototypique (*frame*) : élément de l'énoncé - rôle → représentation sémantique des énoncés
- Déclencheurs / participants / propriétés

Exemple :

Jean **vend** une voiture en Auvergne.

Frame = vente | Déclencheur = «vend»

Participant = « Jean » (vendeur)

Participant = « voiture » (bien vendu)

Propriété = « Auvergne » (lieu)

Typologie

N°	Type	Exemples de marqueurs	Exemples de FA
1	Fuite	fuite, fuyard, écoulement	Fluidic_motion
2	Signal qui s'est déclenché	témoin, alerte, alarme	Warning
3	Obstacle	gêne, empêche, bloque	Hindering
4	Dégradation - Usure - Saleté	cassé, marques, corrosion	Damaging
5	Élément manquant	absence, sans, manque	Presence
6	Configuration hors spécification	hors spécification	Meet_specifications
7	Dispositif qui ne fonctionne pas	HS, panne	Being_operational
8	Action difficile ou impossible	impossible, difficulté	Difficulty
9	État du monde		

Une typologie, pour quoi faire ?

1. Quelle mise en œuvre de la typologie ?

Classification automatique des “déclencheurs”

Hypothèse :

1. Une classe d'expression d'un problème permet de couvrir les variations exprimées dans le corpus.
2. Il existe des marqueurs propres à chaque type.
3. Une classification automatique est possible à partir de l'annotation à gros grains des marqueurs.

Objectifs :

1. Dégager des classes lexicales pour chacune des catégories
2. Première classification des FA en fonction du type d'expression d'un problème et du niveau d'informations présents

Méthode

- **Annotation** : Sélection libre du segment dans le texte et typage sans prétraitements

Exemple d'annotation : FUIITE D'EAU AU NIVEAU DU DISTRIBUTEUR DE COMMANDE DES PLAQUES DE CALAGE DANS LE LOCAL DE LEVAGE

- **Classification** : *Fine-tuning* d'un *transformer* pré-entraîné sur une tâche de classification de *tokens*

Annotation d'un échantillon

Cadre :

- ❑ 3 annotateurs non experts et 1 annotateur expert
- ❑ 50 descriptions d'un problème → validation du guide, consignes d'annotation viables
- ❑ Même corpus afin de comparer les résultats – calculer l'accord inter-annotateur

Consignes :

- ❑ Au moins une annotation par FA
- ❑ Possibilité d'annoter la FA comme incompréhensible
- ❑ Choix de la taille du segment

Annotation - Sélection libre

MESSAGE D'ERREUR LORS DE LA GENERATION DES TABLES IMAGES. LES COEFFICIENTS DES MESURES

M2TK600 ET M2TK604 NE SONT PAS CORRECTS : LES SEGMENTS S1 ET S2 SONT DISJOINTS CF

ANNEXE.

Segment 1

Annotateur 1 :

“MESSAGE”

Annotateurs 2 et 3 :

“MESSAGE D'ERREUR”

Segment 2

Annotateurs 1 et 3 :

“NE SONT PAS CORRECTS”

Annotateur 2 :

“PAS CORRECTS”

Segment 3

Annotateurs 1 et 3 :

“DISJOINTS”

Annotateur 2 :

-

(D=145)
APPARITION D'UN MESSAGE D'INFORMATION DANS LE FICHIER LOG DE L'OUTIL "% CONVERT - I -

DUP.

Annotateur 1 :

“APPARITION D'UN MESSAGE”

Annotateur 2 :

“APPARITION D'UN MESSAGE D'INFORMATION”

Annotateur 3 :

“MESSAGE D'INFORMATION”

Scores d'accord

Annotateurs **naïfs** vs. annotateur **expert** :

Kappa : prise en compte des cas avec recouvrement non-nul des segments (dès que les segments ont un offset en commun).

	Kappa
A1 & A2	0,77
A1 & A3	0,81
A2 & A3	0,82
A1 & Expert	0,58
A2 & Expert	0,67
A3 & Expert	0,80

	Gamma
A1, A2 & A3	0,64
A1, A2 & A3 + expert	0,63

Kappa de Cohen plutôt proche entre les trois annotateurs non-experts.

En revanche, scores plus éloignés avec l'annotateur expert.

Annotateurs naïfs vs. annotateur expert :

Non-experts : **Absence**

Expert : **Fonctionne Pas**

PERTE ◊ DES ◊ 2 ◊ FBF ◊ CCX ◊ T2 ◊ 11 ◊ sec ◊ APRES ◊ LE ◊ DECOLLAGE ◊ (H0) ◊ CF ◊ JDB ◊ JOINT

Non-experts : **Absence**

Expert : **Fonctionne Pas**

NON ◊ ACQUISITION ◊ YSY503.

Non-experts : **Hors Spécifications**

Expert : **Signal**

LE ◊ PARAMETRE ◊ DPRIE ◊ EST ◊ INVALIDE ◊ LORSQUE ◊ PMINRO ◊ OU ◊ PIEOESC ◊ EST ◊ INVALIDE

Classification - Paramètres

Fine-tuning d'un transformer pré-entraîné sur une tâche de classification de *tokens*

Nombre de documents total dans le corpus : 35 242 FA

Nombre de documents annotés : 1000 FA sélectionnées aléatoirement

Répartition par type :

- Très inégale
- Représentative du corpus complet
- Hors Spécification majoritaire et surreprésenté par rapport aux autres

Répartition train / test : 80/20

Pré-traitement : Passage de tout le texte en minuscule

Type	Nombre de segments
Hors Spec	570
Absence	160
Dégradation	144
Fonctionne pas	141
État du monde	102
Impossible	95
Fuite	54
Signal	37
Obstacle	22
Total	1325

Classification - Résultats

Résultats :

Évaluation de différents modèles en précision, rappel et F-mesure :

Modèle	Précision	Rappel	F-mesure (macro-moyenne)
bert-base-multilingual-uncased	0,68	0,68	0,68
bert-base-multilingual-cased	0,65	0,63	0,64
camembert-base	0,63	0,68	0,65
camembert-large	0,65	0,69	0,67
camemberta-base	0,65	0,68	0,66
xlm-roberta-base	0,63	0,65	0,64

Classification - Résultats

Scores pour chaque catégorie :

Métrique *entity-type* : mesure qui prend en compte que l'étiquette du segment et non la justesse des frontières

Score <i>entity-type</i>				
Catégorie	Précision	Rappel	F-mesure	Effectifs
Fuite	0,92	1	0,96	1803
Signal qui s'est déclenché	0,89	0,73	0,80	973
Présence d'un obstacle	0,75	0,60	0,67	526
Dégradation - Usure - Saleté	0,88	0,75	0,81	5559
Élément manquant	0,73	0,86	0,79	5737
Configuration hors spécification	0,68	0,76	0,72	16 188
Dispositif qui ne fonctionne pas	0,73	0,82	0,77	4206
Action difficile ou impossible	0,72	0,86	0,78	2997
État du monde	0,55	0,48	0,51	4297
Total (macro-moyenne)	0,72	0,77	0,74	41 396

- 1) Fuite : résultats particulièrement bons, pas étonnant car peu de marqueurs (fuite, fuyard, écoulement, étanche, étanchéité)
- 2) État du monde : bons résultats pour un type dont le marqueur n'a pas de trait sémantique, il était demandé aux annotateurs de sélectionner le premier token de la FA, → bien repéré par le modèle
- 3) Obstacle : score le plus bas (à l'exception d'État du monde). Étiquette avec le moins de données d'entraînement

Classification automatique des “déclencheurs”

Conclusion :

- Résultats satisfaisants malgré les particularités du corpus et le déséquilibre dans la répartition des catégories au sein du corpus
- Résultats valident la typologie : on a bien des marqueurs propres aux différentes catégories
- Obtention d'une classification du corpus entier (35 242 FA)

Anomalies :

- ❖ Résultats moins performants sans l'étiquette État du Monde
- ❖ Anomalies dans l'annotation BIO en sortie du modèle

Et les *frames* dans tout ça ?

Cas d'étude : la "Fuite"

1. Quelle représentation sémantique d'un dysfonctionnement technique ?
2. Est-il possible d'émettre une grammaire d'un problème technique ?
3. Quels sont les éléments présents et comment sont-ils connectés ?

Pourquoi la fuite ?

Certaines catégories peu adaptées à un traitement par TRIZ en l'état :

- Expression du problème trop “bas niveau”. Ex : “Signal” ou “Hors Spécification”
- Insuffisance d'information exprimée. Ex : “Impossible” ou “Fonctionne Pas”

→ Choix de la catégorie “**Fuite**”

Objectifs :

Représentation sémantique de type *frame* de la fuite dans les FA

Repérer et extraire les éléments nécessaires à la formalisation en vépole

Vépole minimal type pour la Fuite :



Pourquoi la fuite ?

Avantages :

- “Déclencheurs” (marqueurs lexicaux) facilement détectables
- Régularité dans la structure d’expression
- Adéquation avec le *frame* de *Fluidic_motion*. On y trouve les participants pertinents :
 - **Fluid** (ce qui fuit)
 - **Goal** (sur quoi ça fuit)
 - **Source** (d’où ça fuit)
 - **Area** (où la fuite a lieu)

Difficultés :

- Incompréhension des éléments du système sans connaissances expertes
- Qualité d’écriture qui rend impossible une analyse locale fine
- Expert peu disponible

Constats sur la Fuite

1. Impossible de différencier les lieux où se déroule la Fuite

FUITE D'HUILE SUR LE PE DE L'ARBRE DE TRANSMISSION AU NIVEAU DU REDUCTEUR DE L'ASCENSEUR DU BIL.

2. Informations spécifiques non couvertes par le *frame* : quantification de pression / volume / débit, condition d'apparition de la fuite

LORS DE L'ESSAI DE PRESSION DU RESEAU SUITE AU RACCORDEMENT DU RESEAU MASSIF, APPARITION D'UNE FUITE D'EAU AU NIVEAU DE LA SOUDURE EQUIPAGE ALIM TABLE DANS LT CLIM DU MASSIF. **APPARITION DE LA FUITE A UNE PRESSION DE 9 bars.**

3. Éléments difficiles à interpréter ou inconnus

FUITE **AU SNOOP** A L'INTERFACE CVPS02 (PCPH) COMMANDE FERMETURE VPSO.

Approche

Hypothèse :

Il est possible de se baser sur la structure de la phrase pour identifier les composants du *frame*

Approche en trois étapes :

1. Observer les régularités et les récurrences
2. Entretien avec l'expert pour l'explicitation du vocabulaire de base et le sens des structures locales fréquentes
3. Annotation et retours

Étape 1 : *Bleaching*

FA	FA <i>bleached</i>
fuite sur le process de maintien en fermeture virh gam conduisant a regonfler la bouteille toutes les 24 heures	fuite sur le X de X en X a X la X le NUM X
fuite au snoop a l' interface cvps02 (pcph) commande fermeture vps0 .	fuite à le X a l' X NUM X

Se base sur les structures de phrases et notamment les prépositions

Remplacement de :

- mots pleins par X
- termes avec une fréquence < 60 occurrences par **X**
- chiffres ou code avec des chiffres par **NUM**
- fluides par **LIQUIDE** ou **GAZ**

Étape1 : analyseur syntaxique

Bruit + langage spécialisé = performances d'un analyseur syntaxique dégradées

Objectif : Extraction de *patterns* récurrents et observation de leur usage en contexte

Méthode :

Application de Stanza (Qi et al., 2020)

Extraction de patterns

Étape 1 :Analyse syntaxique

```
# text = fuite sur raccord des liaisons eccn1 ( pcpo ) et virh ( pcpo ) detectees au snoop .
# sent_id = 135
1 fuite fuite NOUN _ _ Gender=Fem|Number=Sing 0 root _ start_char=22420|end_char=22425
2 sur sur ADP _ _ 3 case _ start_char=22426|end_char=22429
3 raccord raccord NOUN _ _ Gender=Masc|Number=Sing 1 nmod _ start_char=22430|end_char=22437
4-5 des _ _ _ _ start_char=22438|end_char=22441
4 de de ADP _ _ 6 case _ _
5 les le DET _ _ Definite=Def|Number=Plur|PronType=Art 6 det _ _
6 liaisons liaison NOUN _ _ Gender=Fem|Number=Plur 1 nmod _ start_char=22442|end_char=22450
7 eccn1 eccn1 X _ _ 6 appos _ start_char=22451|end_char=22456
8 ( ( PUNCT _ _ 9 punct _ start_char=22457|end_char=22458
9 pcpo pcpo X _ _ 7 appos _ start_char=22459|end_char=22463
10 ) ) PUNCT _ _ 9 punct _ start_char=22464|end_char=22465
11 et et CCONJ _ _ 12 cc _ start_char=22466|end_char=22468
12 virh virh X _ _ 7 conj _ start_char=22469|end_char=22473
13 ( ( PUNCT _ _ 14 punct _ start_char=22474|end_char=22475
14 pcpo pcpo X _ _ 12 appos _ start_char=22476|end_char=22480
15 ) ) PUNCT _ _ 14 punct _ start_char=22481|end_char=22482
16 detectees detectees X _ _ 12 appos _ start_char=22483|end_char=22492
17-18 au _ _ _ _ start_char=22493|end_char=22495
17 à à ADP _ _ 19 case _ _
18 le le DET _ _ Definite=Def|Gender=Masc|Number=Sing|PronType=Art 19 det _ _
19 snoop snoop NOUN _ _ Gender=Masc|Number=Sing 16 nmod _ start_char=22496|end_char=22501
20 . . PUNCT _ _ 1 punct _ start_char=22502|end_char=22503|SpacesAfter=\n\n
```

fuite sur raccord

fuite (detectee à le snoop) **sur** le **raccord** entre le flexible epsrh et la pcp .

fuite a 10 - ² **sur** **raccord** ombilical de purge h2 esc .

fuite sur raccord de les liaisons eccn1 (pcpo) et virh (pcpo) detectees à le snoop .

fuite d' huile **sur** un **raccord** de le verin de levage c durant l' operation de levage .

Extraction des *patterns* Fuite (gouverneur) + PREP+ NOM

Étape 2 : Entretien avec l'expert

Définition du vocabulaire fréquent :

- Table : table de lancement
- LH2 : Hydrogène liquide
- bulles : moyen de détecter une fuite
- ...

Sens et typage de structures fréquentes :

- “au snoop” : liquide détecteur de fuite
- “en ligne” ; indication que la fuite est interne
- ...

Passage en revue des structures avec prépositions : différenciation entre sur / au / au niveau de...

Frame de la Fuite dans un environnement technique

Élaboration d'un *frame* de la Fuite dans un environnement technique :

Participants	Définitions
Fluide	Liquide ou gaz qui fuit
Dispositif de détection	Dispositif utilisé pour détecter la fuite
Débit	Débit hors standard (conséquence ou signal de la fuite)
Quantité	Renvoie à la quantité de fluide (ex : fuite importante, légère...)
Élément du processus fluide	Élément précis du processus fluide (ex : "joint", "vanne")
Circuit ou système	Circuit ou système avec une composante hydraulique (ex : "partie eau froide du compresseur")
Localisation géographique	Localisation géographique générale (ex : "Bâtiment d'assemblage")

Frame Element
Fluide
Dispositif de détection
Débit
Quantité
Localisation précise
Localisation moyenne
Localisation géographique large

LEGERE FUIITE D'HUILE AU NIVEAU DE L'ECROU INFERIEUR DE L'INDICATEUR DE NIVEAU DU SKID
GAM/GAT

Étape 3 : Annotation et retour

Annotation fine d'un échantillon de 50 FA par l'expert

→ Identification d'un élément maquant au *frame* :

“Condition d'apparition d'une fuite”

Étiquette	Fréquence
Localisation précise	40
Localisation moyenne	34
Dispositif de détection	13
Débit	12
Fluide	11
Localisation géographique large	8
Quantité	6
Condition	2
Total	124

Fuite : Articulation avec TRIZ

Exemple

LEGERE FUIITE D'HUILE AU NIVEAU DE L'ECROU INFERIEUR DE L'INDICATEUR DE NIVEAU DU SKID GAM/GAT

Étape 1

LEGERE **FUIITE** D'HUILE AU NIVEAU DE L'ECROU INFERIEUR DE L'INDICATEUR DE NIVEAU DU SKID GAM/GAT → **Type : Fuite**

Étape 2

LEGERE FUIITE **D'HUILE** AU NIVEAU DE L'ECROU INFERIEUR DE L'INDICATEUR DE NIVEAU DU SKID GAM/GAT → **Identification des participants**

Si les éléments **Élément du process fluide** + **Fluide** sont présents dans le texte
→ obtention des éléments minimaux pour un **vépole** :

Élément du process fluide - **NE RETIENT PAS** - **Fluide**

Les champs “Cause” et “Action Réalisée”

Sur le corpus complet

Cause :

- Pas systématiquement renseignés (30% manquant)
- Diagnostique du problème
- Moyenne de 5,4 mots par FA (contre 19,3 dans la description du problème)

Action réalisée :

- Renseignée dans la majorité des cas (2,7% manquant)
- 8,4% de “EN L’ÉTAT” (“Acceptable en l’état”, “Vie en l’état”) pour signaler qu’aucun changement n’a été effectué
- Moyenne de 12,7 mots par FA

Fuite - “Cause” et “Action réalisée”

Constat :

Les causes et les actions réalisées présentent moins de variations dans les énoncés.

Objectif :

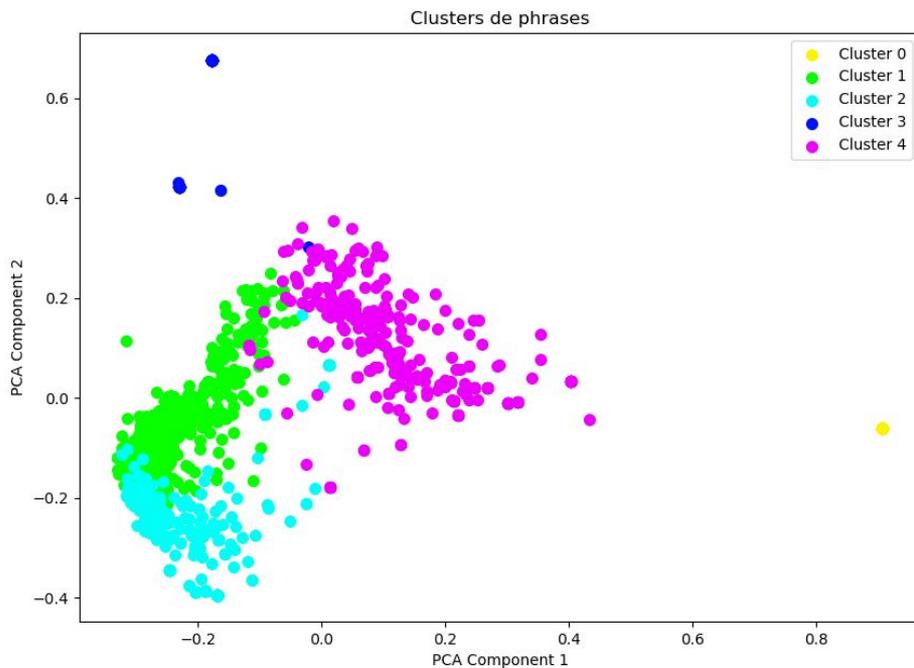
- Regrouper les causes similaires.
- Regrouper les actions réalisées similaires.

Méthode :

Utilisation d'un LLM pour produire des *embeddings* (plongement de mots) de FA

Clustering (distance de cosinus).

Clusters - Cause



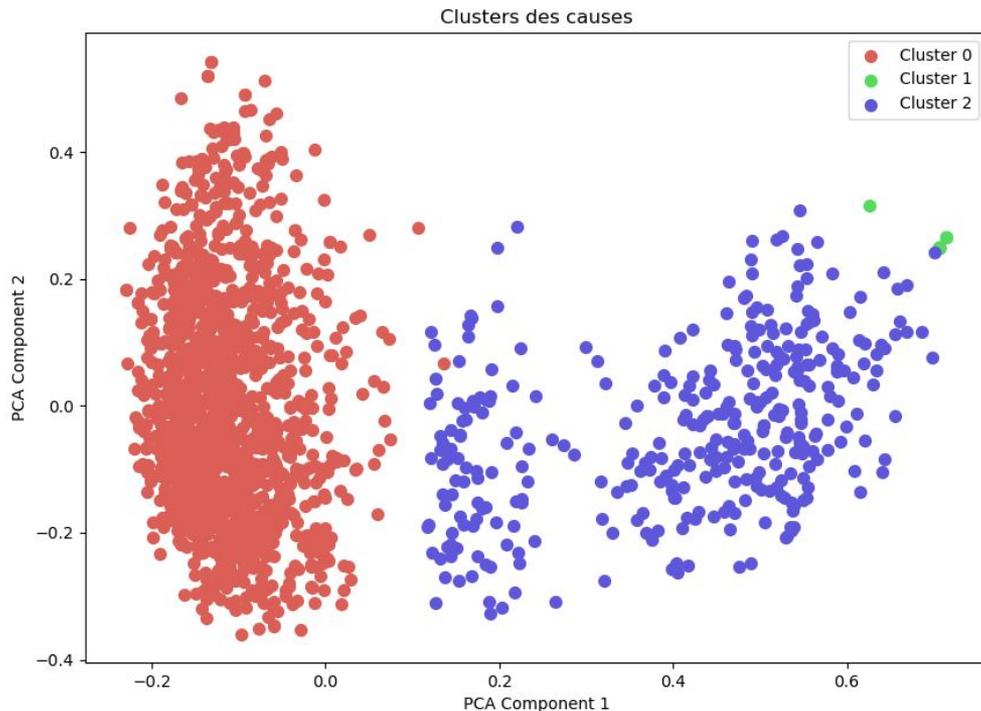
Bilan mitigé :

- On identifie deux clusters bien distincts (**0** et **3**) qui correspondent respectivement à un énoncé :
“NEANT” ou “En cours...”
- Mais “NEANT.” se retrouve dans le cluster 3. À première vue :
 - 2 à 3 mots + PUNCT : cluster 2
 - > 3 mots + PUNCT : cluster 1
 - pas de PUNCT (hors “NEANT”) : cluster 4

Clusters - Cause

0	1	2	3	4
NEANT	BILLE DEFECTUEUSE.	*	En cours ...	FLAMME ESC-124 EN DOUBLE EXEMPLAIRE. IL Y AVAIT 2 FLAMMES A LA LIVRAISON
NEANT	1 - MAUVAIS MONTAGE DU JOINT EN CUIVRE A L'INTERFACE SOL/BORD. 2 - MAUVAIS MONTAGE.	*	En cours ...	LIE A LA DEFINITION
NEANT	LIGNE AD (FERMETURE CLA33 PAÇ O2) DEFORMEE, TUYAUTERIE TORDUE. SA GEOMETRIE/FORME NE CORRESPOND PAS A L'ETAT ATTENDU.	*	En cours ...	ERREUR OPERATEUR COTE SOL LORS DES RACCORDEMENTS COTE MAT
NEANT	1- NON CONFORMITE2- CAUSE PROBABLE : FUI TE DU POINTEAU (VIRH CONTROLEE FERMEE A J6) 3 - POSITIONNEMENT DE LOT DEPORTE DU TEGOD.	*	En cours ...	HORS SPECIFICATION
NEANT	CTRL DE SURFACE SUR LIAISON SOL.	DEGRADATION	En cours ...	NON CONFORMITE
NEANT	HORS FAMILLES3- 1 PIQUAGE CIRCUIT DE VENTIL PTM LAISSE NON OBTURE (J DE MEO)4- CE37 : FUITES IMPORTANTES CHAMBRE + DIV2- SOMME DES FUITES EYSCR ENTRE AUTRES.	SIEGES CLAPETS MARQUES.	En cours ...	LIVRAISON
NEANT	EXIGENCE TRES SERREE PAR RAPPORT A CE RACCORDEMENT.	CLAPET NON ETANCHE.	En cours ...	CLAPET MARQUE ?
NEANT	MAUVAIS SERRAGE INITIAL.	*	En cours	DEGRADATION JOINT INTERNE
NEANT	CONFIGURATION DU TEST.	*	En cours	FUITES
NEANT	CONFIGURATION D'ESSAI + MOYEN DE MESURE. PROBABLE : FUI TE PARASITE AUX I/F SOL/BORD OU SOL/SOL.	LANCEMENT.	En cours	PROBABLEMENT DUE AU VIEILLISSEMENT, JOINTS DETERIORES SUREMENT LORS DE LEUR MEP, PAS D'OUTILS APPROPRIES
NEANT	MAUVAIS SERRAGE I/F CAPA.	DEGRADATION PORTEE.	En cours	DEGRADATION MATERIEL
NEANT	FUI TE DANS LE TOIT DU BAF. ZONE SAS.	* RACCORD DESSERRE.	En cours ...	DEGRADATION PAR UTILISATION NORMALE PROBABLE : MANQUE RONDELLE
NEANT	TAUX HELIUM AMBIANT ELEVE (+ ZONE CONFINEE)	DEGRADATION.	En cours	DEGRADATION. OXYDATION DU FOND DE CUVE
NEANT	AMBIANCE SATUREE. FUI TE RESIDUELLE.	GARNITURE MECANIQUE HS.	En cours	DEGRADATION FLEXIBLE PAR UTILISATION NORMALE PROBABLE :1) PROBLEME MECANIQUE DETENDEUR -> RESSORT DEFECTUEUX2) PROBLEME ETANCHEITE CLAPET -> POUSSIERE/SIEGE DU CLAPET
NEANT	CAUSE INDETERMINEE, ANTERIORITES : B164/116 - B162/46 - B142/49.	VIEILLISSEMENT.	En cours	VIEILLISSEMENT
NEANT	NON CONFORMITE INDUSTRIELLE.	FLEXIBLES FUYARDS PROBABLE.	En cours	VIEILLISSEMENT
NEANT	NON CONFORMITE INDUSTRIELLE.	NON DETERMINEE.	En cours	UTILISATION CAMPAGNE 502
NEANT	FUI TE SUR OBTURATEUR.	CLAPET FEMÀ HS.	En cours	DEGRADATION ETANCHEITE DE LA VANNE
NEANT	1)MALADRESSE2)Pb. D'INTERFERENCE.	SERRAGE COLLIER MOU CORRECT.	En cours	DEGRADATION PRESSE ETOUPE
NEANT	REALISATION LIAISON RIVETEE EN USINE.	* VANNE FUYARDE.	En cours	DEGRADATION VANNE
NEANT	CONDITIONS DE TEST.		En cours	JOINT TORIQUE HORS D'USAGE
NEANT	ANTERIORITES : B128/62 - B130/26 - B135/48 - B140/29 - B161/23 - B162/34CAUSE : NON DETERMINEE.	VIEILLISSEMENT.	En cours	JOINT HS
NEANT	A CERTAINS REGLAGES, LE DETENDEUR CLIENT EST FUYARD.	*	En cours	PAS DE PROBLEME SUR LE FLEXIBLEPROBLEME DE SERRAGE ENTRE FLEXIBLE ET FILTRE 5 MICRONS
NEANT	GESTION DE CONF. ANOMALIE CONFIGURATION.	NEANT.	En cours	NOVINTECCHANGEMENT DU JOINT ROSAN
NEANT	ANTERIORITES: B164/116 (ECCN) - B167/75 - B168/64 (CVIRH).	NEANT.	En cours	ERREUR D'INTERPRETATION DES RESULTATS
NEANT	AMBIANCE POLLUEE > 10-5 ANTERIORITE : B161/89.	NEANT.	En cours	FUI TE EN LIGNE
NEANT	DEFAILLANCE FONCTIONNELLE.	NEANT.	En cours	PRESSE ETOUPE DESSERE
NEANT	LA ZONE DE PERÇAGE EST D'UN DIAMETRE AVOISINANT LE DIAMETRE DE DEUX ALESAGES (DIAM 18).	NEANT.	En cours	VIEILLISSEMENT
NEANT		VIBRATION.	En cours	JOINT HS
NEANT	DIFFICULTES D'ACCOSTAGE.	VIEILLISSEMENT.	En cours	PROBLEME DETANCHEITE
NEANT	CONCEPTION DE L'ADAPTATEUR.	VIEILLISSEMENT.	En cours	JOINT TEFLON FENDU
NEANT	ATMOSPHERE ENVIRONNANTE.	AD.	En cours	VIEILLISSEMENT
NEANT	PETITES DEGRADATIONS - ERREUR DE PRO.	FILTRE FUYARD.	En cours	DESSERRAGE POSSIBLE DU RACCORD AU MOMENT DU LANCEMENT

Clusters - Action réalisée



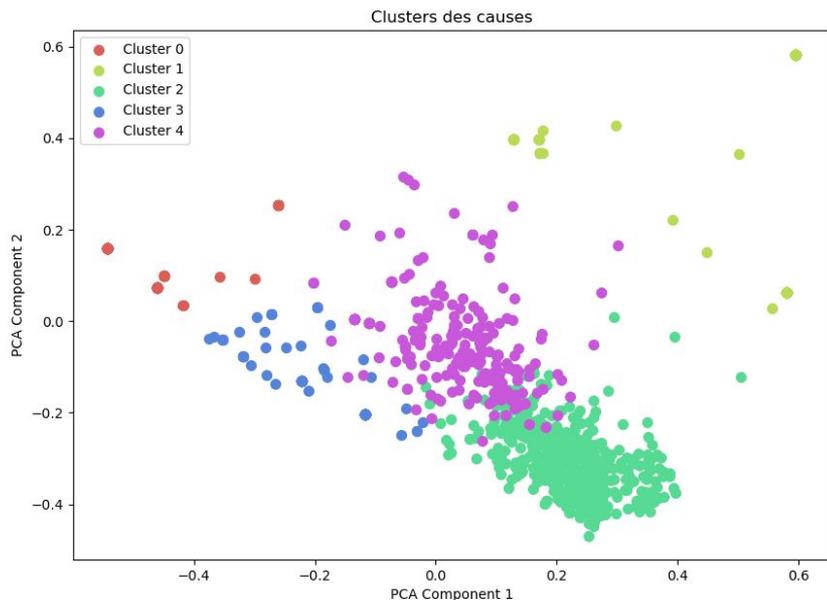
Bilan mitigé :

- **Cluster 1** : occurrences de “nan” et “*”
- **Cluster 2 et 3** : énoncés similaires retrouvés dans les deux clusters avec pour seule différence un “.”

Clusters - Action réalisée

0	1	2
1)OPERATION MS003WW LE 13/05/98 DE 14H A 16H SUIVANT PROCEDURE A5.3 POA112 0250 PO (0) - 7 4.2.2.4 RAS : P COMMANDE (M5) = 73s-2)OPE MA002E0 LE 14/5/98 : CTRL P COMMANDE (M5) = 68s. INVESTIGATIONS : CHUTE DE PRESSION DE 14b (60 A 46b) APRES DECONNEXION VIRH EN 12H. TEST AU SNOOP REALISE : RAS. VOIR NOTE AS CI-JOINTE A1396/98	*	ACCEPTABLE EN L'ETAT
A/ UTILISATION EN L'ETAT/ NETTOYAGE : PROPOSITION : SI CYCLAGE DEMANDE, MISE EN PLACE OUTILLAGE LEFP2 AVEC VANNE DE PURGE.	080	ACCEPTÉ EN L'ETAT
1/ ACCEPTABLE EN L'ETAT/ VERIFICATION DU COUPLE MINI DE SERRAGE DU CAPTEUR/3/ ACCEPTABLE EN L'ETAT/ VERIFICATION PAR CAPILLAIRE - ACCEPTABLE EN L'ETAT).	080	DEPRESSURISATION EPSRRH A 0.277 BAR REL DEPRESSU CVRRHREPRESSURAGE A 17.5N.m DES DELUX LIAISONSPRESSURISATION DU 70 Bar ET DE LA LIGNE EPSRHCONTROLE ETANCHEITE CVRRH TOUJOURS A 1.10-3 Ncm3/s ET EPSRH A 2.10-3Ncm3/sACCEPTABLE PAR SMOG
CONTROLE D'ETANCHEITE (JOINT GONFLE SUR LES PPM)FAIT. FUIITE 0.SI OK REPRISE BALAYAGE DES LIGNE PPM (SINON ARRET DE L'OP ET CHANGEMENT DU JOINT)PAS DE REPRISE DES TENEURS CHAMBRE.	*	RAS
NON REALISATION DU DETROMPAGE CVRH DANS L'OPERATION ET NON FERMETURE DE LA VHP (NOTA : IL S'AGIT D'UN CAS PREVU EN PROCEDURE). LE DETROMPAGE CVRH ET LA FERMETURE VHP SERONT REALISES DANS L'OPERATION MC208E0. ETANCHEITE POP APRES RACCORDEMENT EPC. (CAS DEGRADEES PRIS EN COMPTE DANS LES PROCEDURES).L'OBTURATEUR EST REMIS EN PLACE SUR LA FUIITE CALIBREE PGD.	*	ACCEPTABLE EN L'ETAT (VOIR NOTE D0).
ACCEPTÉ EN L'ETAT.	080	CONFORME A PROPOSITION RT
ACCEPTABLE VOIR NOTE AE/DI NR 088/99.	080	REMPLACEMENT DU CLAPET N°98621297 PAR LE N° 98621499 DANS L'OPERATION WM501 E0 (114) AVEC PROCEDURE SPECIFIQUE : A53 POA0140 2163 C0 (0).TEST ETANCHEITE CORRECT.DEMANDE D'EXPERTISE SUR CLAPET RENVOYE EN METROPOLE (AE/DI/AT 140 5801).
VOIR NOTE AE/DI NR 91/99. ACCEPTÉ EN L'ETAT.	080	NEANT
CONTROLE CORRECT APRES REMPLACEMENT DES OBTURATEURS.	080	VIE EN L'ETAT. VOIR NOTE AE/DI/AT V153/N°69/02
ACCEPTÉ EN L'ETAT. VOIR NOTE JOINTE AE/DI/AT 101/99. ACTION DE NIVEAU 2 PRISE EN COMPTE.	080	ACCEPTABLE SUIVANT NOTE AE/DI/AT/V157/2002-129 DU 07/10/02.19/10/02 - VOIR SMOS 994a
PLUS DE FUITES DURANT LE REMPLISSAGE. AUCUN IMPACT SUR LE VOLUME INTROUIT DANS LE GAM (BORD NON RACCORDE). EQUIPEMENT SOL REVALIDE.	080	- TEST GLOBAL CONFORME-LA FUIITE EST NULLE COTE SOL-TEST A REPRIRE LORS DES CTRL ETANCHEITE 2EME PARTIE (VMFO ACTIVEE ENTRE TEMPS). DEPOSE/REPOSE LIGNE 132 : PROC : A53 POA158 2512 (WMS50650).ETANCHEITE REALISEE CORRECTE : PROC : A53 POA158 2513 (WMS50750). 14/05/03 (J12) OPE AJOURNEE SUITE FA B161/050 => J16.ETANCHEITE REPRISE EN PASSAGE DE LA YMCJ09E "ETANCHEITE MOTEUR VULCAIN APRES CHANGEMENT DIVERGENT".
DEJA VU EN INTEGRATION CHEZ MATRA.	080	DERACCORDEMENT LIGNE : CT I/F RAS. RERACCORDEMENT AU COUPLE + 10%: FUIITE PERSISTANTE AU SNOOP.REDERACCORDEMENT LIGNE. TOILLAGE OLIVE. REMONTAGE CT SNOOP OK
MISE EN PLACE DE SCOTCH KARTON POUR ETANCHER LE FLEXIBLE.	080	
RAS.	080	
-DEPOSE DE LA LIGNE 104 DANS LA WMS503S0 (2511. J11)-EXPERTISE DES PIECES.-DEPOSE DE LA LIGNE 104 DANS LA WMS505S0 (2515. J14). ETANCHEITE CORRECTE DANS LA WMS505S0 (J16) CORRECTE.+ RECOMMANDATION PROCEDURE ET SMO.	080	VIE EN L'ETAT POUR RSL2. SEFF ENVIRON 20.10cm3 EN PRENANT EN COMPTE LE POINT MILIEU DE L'OFFSET DU DEBITMETRE (CF FA B164/424).
RAS.	080	ACCEPTABLE EN L'ETAT
- SERRAGE AU COUPLE 17.5 N.m DES 2 LIAISONS CI-DESSUS. FUIITE RELEVÉE COMMANDE ECROU/MAMELON: 1.10E-5 NcmE3/s FUIITE RELEVÉE ALIMENTATION ECROU/MAMELON : 5.10E-3 NcmE3/s.VIE EN L'ETAT.	080	FAIT DANS L'OP MEME : RASMESURE DE LA FUIITE : 2.3.10-4Ncm3/s
RAS.	080	LE COUPLE MAXI A ETE PASSE SUR LES 2 LIAISONS : SUR CHACUNE. FUITES RESIDUELLES (MICROBULLES). TEST REALISE AVEC POCHON : FUIITE < 10-3. (AMBIANCE ENTRE 2 ET 8.10-4)
ACCEPTÉ EN L'ETAT.	080	

Clusters - Cause sans ponctuation finale



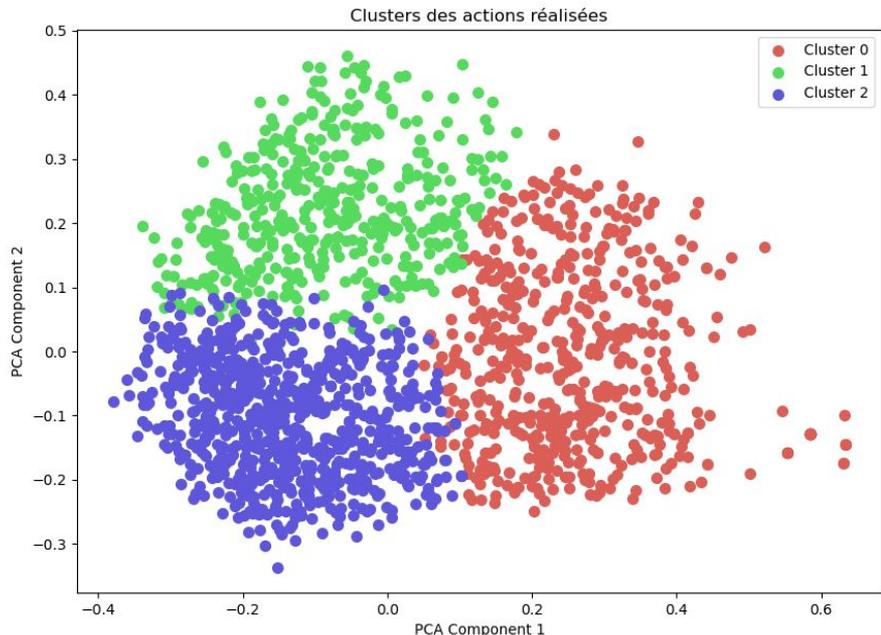
Bilan :

- **Cluster 0** : un seul mot → “NEANT” ou “FUITE” ou “USURE”
- **Cluster 1** : “En cours...” ou “*” ou “AD”
- **Cluster 2** : Énoncés plutôt longs en majorité, mais pas que
- **Cluster 3** : un à deux mots → “VIEILLISSEMENT”, “CORROSION”, “DEGRADATION”...
- **Cluster 4** : Tous les énoncés tendant vers “NON DETERMINEE” mais pas que. La majorité sont plutôt courts

Clusters - Cause sans ponctuation finale

0	1	2	3	4
NEANT	AD	PROBABLE : FUI TE AU SEIN DE L'ARMOIRE	CONTRUCTION	A DEFINIR
NEANT	En cours ...	PROBABLE : SYSTEME DE MAINTIEN DES FLEXIBLES INSUFFISANT. MISE EN CONTRAINTE	CONSTRUCTION	DEFAULT INDUSTRIEL
NEANT	En cours ...	A/ OUTILLAGE NON DEPOSE A LA FIN DU CYCLAGE/ PJEAGE D'HUILE DANS LIGNE LEFP2 A LA SUITE DU CYCLAGE + ABSENCE DE LA VANNE DE PURGE	AMBIANCE	INDETERMINEE
NEANT	*	1/ PROBLEME DEFINITION DE FUI TE INTERNE LV CONNUE ET PREVUE EN SMOS/CAS DEGRADE,2/ INCONNUE - NON REPEREE EN RECETTE CASE3/ FUI TE STABLE REPEREE ET ACCEPTEE EN RECETTE CASE MAIS LEGEREMENT INFERIEURES A LA SPEC,4/ IDEM	DEFAILLANCE	INDETERMINEE
NEANT	En cours	PROBABLE FUI TE DE L'AIR DE GONFLAGE DU JOINT VERS LA CHAMBRE	LIVRAISON	INDETERMINEE
NEANT	En cours	LE SYSCOM HP DE L'EPC EST DEPRESSURISE, (NON CONFORMITE A LA CONFIGURATION DE TRANSPORT ETAGE)	DEGRADATION	MAINTENANCE INADAPTEE
NEANT	En cours	NON CONFORMITE	FUITES	MAINTENANCE INADAPTEE
PROCEDURE IMPRECISE	En cours ...	VOIR NOTE JOINTE AE/DI/AT 101/99	VIIEILLISSEMENT	INDETERMINEES
NDT	*	FLEXIBLE PROBABLEMENT GRIFFE LORS DE SA MISE EN PLACE	VIIEILLISSEMENT	EN COURS
NDT	*	PROBABLES :- ABSENCE SUPRANITE,- COUPLE NOMINALE APPLIQUE : 8,8 AU LIEU DE 8 (MAXI DE LA PSEC.) -> PERTE D'EFFICACITE DU SYSTEME DETANCHEITE	VEILLISSEMENT	NON DETERMINEE
NEANT	En cours ...	NON DETERMINEE (LANCEUR)	LANCEMENT	NON DETERMINEE
NDT	*	NON DETERMINEE (LANCEUR)	DEGRADATION	CAPTEUR PRB NOYE DANS LE FHS
FUI TE	*	NON DETERMINEE (LANCEUR)	VIIEILLISSEMENT	ANTERIORITES : NON
FUI TE	En cours ...	LE SYSCOM HP DE L'EPC EST DEPRESSURISE	VIIEILLISSEMENT	FUI TE CLAPET CEVC AU EV
NEANT	En cours ...	DETENDEUR PRESQUE FERME. (V7 OUVERTE, V6 ET V5 OUVERTES)	VIBRATION	TEMPERATURE TROP BASSE DANS LA JAVE EPC
NEANT	En cours...	ERREUR OPERATEUR	VIIEILLISSEMENT	CONCEPTION NON ADAPTEE A LA SPEC
NEANT	*	MAUVAISE LIAISON	VIIEILLISSEMENT	JOINT DEFAILLANT
NEANT	*	SPECIF TROP SERREE (PM EN COURS POUR ELOIGNEMENT DU CRITERE)	VIIEILLISSEMENT	FUI TE REMARQUEE AU SNOOP AU NIVEAU DU JOINT GONFLABLE DU TQD
NEANT	En cours ...	ACCOSTAGE EVENTUELLEMENT DEGRADE PAR REPRISE MECANIQUE SUPPORTAGE	VIBRATION	CAUSE INDETERMINEE
NEANT	En cours ...	NON CONFORMITE INDUSTRIELLE. ANTERIORITES: OUI - FA B135/1	VIIEILLISSEMENT	INDETERMINEE
NEANT	En cours ...	NON DETERMINEE. ANTERIORITES : OUI -> FA B142/39 + B140/29	CORROSION	MICRO FUI TE COMMANDE VCOF
FUI TE	En cours ...	1 - MESURES DECALEES,2 - JOINT ENDOMMAGE,ANTERIORITE : NON	VIIEILLISSEMENT	DEFAILLANCE OUTILLAGE SOL
NDT	En cours ...	ANTERIORITES : OUI -> B128/62 - B130/26 - B135/48 - B140/29 - B142/39 - B145/72	VIIEILLISSEMENT	INDETERMINEE
NEANT	En cours ...	ANTERIORITES : NON LE TUBE DE LA COLONNE EST CASSE AU NIVEAU DE LA GRADUATION 120	VIIEILLISSEMENT	NON DETERMINEE
NEANT	En cours ...	ERREUR MONTAGE	VIIEILLISSEMENT	1- NON CONFORMITE2 - CAUSE PROBABLE : FUI TE DU POINTEAU (VRH CONTROLEE FERMEE A J6).3 - POSITIONNEMENT DE LOT DEPORTE DU TEGOD
NEANT	En cours ...	ANTERIORITES : OUI - FA B119/14 - B128/62 - B153/47,1/ IL MANQUE LE PERCAGE DANS LES RACCORDS POUR PASSER LE FIL FREIN	VIIEILLISSEMENT	CTRL DE SURFACE SUR LIAISON SOL
NEANT	En cours ...	DETERIORATION DUE A LA SURPRESSION MOTEUR A 10 bars (FA 40)	VIIEILLISSEMENT	CAUSE NON DETERMINEE
NEANT	En cours ...	BOUCHON DE PURGE DU TE BP NON SERRE AU COUPLE DE 7,5 Nm	VIIEILLISSEMENT/USURE	FUI TE PAC O2
NEANT	En cours ...	- ABSENCE DEPAULEMENT DE CENTRAGE SUR LES OMBILICAUX SOL.- LE JOINT UTILISE SEMBLE TROP PETIT	VIIEILLISSEMENT	NON DETERMINEE

Clusters - Action réalisée sans ponctuation finale



Bilan :

On retrouve les cas “ACCEPTABLE EN L’ETAT” dans le cluster 2 quand il est seul. S’il est suivi d’autres mots, il peut être retrouvé dans les autres clusters. À première vue, pas de régularités entre les FA d’un même cluster. Le lexique récurrent (“REMPACEMENT”, “ETANCHEITE”, “JOINT”...) est présent dans plusieurs clusters.

Clusters - Action réalisée sans ponctuation finale

1	0	1	2
<p>1)OPERATION MS003MW LE 13/05/98 DE 14H A 16H SUIVANT PROCEDURE AS.3 POA112 0250 PQ (0) - ? 4.2.4 RAS : P COMMANDE (M5) = 73b.2)OPE MA002EO LE 14/5/98 : CTRL P COMMANDE (M5) = 68b. INVESTIGATIONS : CHUTE DE PRESSION DE 14b (80 A 45b) APRES DECONNEXION VIRH EN 12H. TEST AU SNOOP REALISE : RAS. VOIR NOTE AS CI-JOINTE A1396/98</p>	<p>1) ACCEPTABLE EN LETAT/ VERIFICATION DU COUPLE MINI DE SERRAGE DU CAPTEUR3/ ACCEPTABLE EN LETAT4/ VERIFICATION PAR CAPILLAIRE - ACCEPTABLE EN LETAT</p>	<p>ACCEPTABLE EN LETAT</p>	
<p>NON REALISATION DU DETROMPAGE CVHP DANS L'OPERATION ET NON FERMETURE DE LA VHP (NOTA : IL S'AGIT D'UN CAS PREVU EN PROCEDURE). LE DETROMPAGE CVHP ET LA FERMETURE VHP SERONT REALISES DANS L'OPERATION MC208EO ETANCHEITE POP APRES RACCORDEMENT EPC (CAS DEGRADEES PRIS EN COMPTE DANS LES PROCEDURES). L'OBTURATEUR EST REMIS EN PLACE SUR LA FUITE CAUBREE PGD</p>	<p>ACCEPTABLE EN LETAT (VOIR NOTE D) REMPLACEMENT DU CLAPET N°98621297 PAR LE N° 98621498 DANS L'OPERATION WMG501 EQ (J14) AVEC PROCEDURE SPECIFIQUE : A53 POA0140 2163 C0 (0). TEST ETANCHEITE CORRECTE. DEMANDE D'EXPERTISE SUR CLAPET RENVOYE EN METROPOLE (AE/DI/AT 140 5801)</p>	<p>A/ UTILISATION EN LETAT/ NETTOYAGE : PROPOSITION : SI CYCLAGE DEMANDE, MISE EN PLACE OUTILLAGE LEFP2 AVEC VANNE DE PURGE</p>	
<p>ACCEPTABLE VOIR NOTE AE/DI NR 988/99 DEJA VU EN INTEGRATION CHEZ MATRA</p> <p>DEPRESSURISATION EPSRH A 0.277 BAR REL. DEPRESSU CVHP/RESSERAGE A 17.5N.m DES DEUX LIAISONS/RESSERAGE DU 70 BAR ET DE LA LIGNE EPSRH/CONTROLE ETANCHEITE CVHPH TOUJOURS A 1.10-3 Ncm3/s ET EPSRH A 2.10-3Ncm3/sACCEPTABLE PAR SMOG</p>	<p>VIE EN LETAT, VOIR NOTE AE/DI/AT V153/N°69/02 NETTOYAGE DE LA TUYERE (A SEC) DANS L'OPE WEG502P2 A J19 + CHANGEMENT DU POCHON. INSPECTION DES RACCORDS BP DEVAQUATION (OPE A WEG502P2 A J25) ET RESSERRAGE AU COUPLE DU BOUCHON DE PURGE DU TE BP (COUPLE 7.5 Nm) AVEC PROCEDURE SPECIFIQUE 1291 C0 (0) + NOTE AE/DI/AT/V157-2002/71 DU 19/9/02</p>	<p>ACCEPTÉ EN LETAT</p>	
<p>RAS -DEPOSE DE LA LIGNE 104 DANS LA WMS50350 (2511, J11). EXPERTISE DES PIECES, DEPOSE DE LA LIGNE 104 DANS LA WMS5050 (2515, J14) ETANCHEITE CORRECTE DANS LA WMS5050 (J16) CORRECTE. + RECOMMANDATION PROCEDURE ET SMO</p>	<p>2/10/02 : PR CADUQUE I/F SOL BUL EN CAUSE30/10/02 : OUVRIER FA SOL S157/318 SUR LES 2 CAUSES PROBABLES</p>	<p>VOIR NOTE AE/DI NR 91/99. ACCEPTÉ EN LETAT</p>	
<p>RAS -HEU TEST REFAIT : VALEUR 2.10-5. VIE EN LETAT.2.3.4- VIE EN LETAT 3- VERIFICATION DU TRAIT BIERRE ET FREINAGE : OK</p>	<p>1- TEST GLOBAL CONFORME LA FUITE EST NULLE COTE SOL- TEST A REPENDRE LORS DES CTRL ETANCHEITE 2EME PARTIE (VMFO ACTIVEE ENTRE TEMPS)</p>	<p>ACCEPTÉ EN LETAT SUIVANT NOTE AE/DI/AT/V157-2002/218</p>	
<p>RAS -FIN DE PROCEDURE POUR CAS CAS PREVU D'ABSENCE FUITE PGD - RERESSU CARACTE HP A 5 bar. - CTRL HYDRO : 240 ppm</p>	<p>DEPOSE/REPOSE LIGNE 132 : PROC : A53 POA158 2512 (WMS50650) ETANCHEITE REALISEE CORRECTE : PROC : A53 POA158 2513 (WMS50750)</p>	<p>ACCEPTÉ EN LETAT</p>	
<p>RAS 1- NETTOYAGE DE L'HUILE SUR CLAPET BP SERV. VERIN V WEG508P1 LE 26/1/01. 2- INSPECTION APRES CTRL DE SYNTHESE, NOTE AT N°187/04 ACCEPTÉE EN LETAT. INSPECTION AVANT MEP PAR.3- RELEVÉ DE LA COTE LE 9/2/04 : 122.4mm. PRESENCE D'HUILE LE 9/2/04 (ESSUYEE)WEG506P1. NOTE AT N°275/04 SVE ACCEPTABLE EN LETAT LE 18/2/04. RELEVÉ COTE 121.5mm. WEG508P1</p>	<p>1- NETTOYAGE DE L'HUILE SUR CLAPET BP SERV. VERIN V WEG508P1 LE 26/1/01. 2- INSPECTION APRES CTRL DE SYNTHESE, NOTE AT N°187/04 ACCEPTÉE EN LETAT. INSPECTION AVANT MEP PAR.3- RELEVÉ DE LA COTE LE 9/2/04 : 122.4mm. PRESENCE D'HUILE LE 9/2/04 (ESSUYEE)WEG506P1. NOTE AT N°275/04 SVE ACCEPTABLE EN LETAT LE 18/2/04. RELEVÉ COTE 121.5mm. WEG508P1</p>	<p>MISE EN PLACE DE SCOTCH KAPTON POUR ETANCHEER LE FLEXIBLE</p>	
<p>ENVOI NOUVEAU HDI PAR SAAB ET REPRISE SANGLAGE LE 13/10/00 AU COURS DE YMH41486</p>	<p>14/05/03 (J12) OPE AJOURNEE SUITE FA B16/050 => J16. ETANCHEITE REPRISE EN PASSAGE DE LA YMCJ09EO "ETANCHEITE MOTEUR VULCAIN APRES CHANGEMENT DIVERGENT"</p>	<p>- SERRAGE AU COUPLE 17.5 Nm DES 2 LIAISONS CI-DESSUS. FUITE RELEVÉE COMMANDE ECROU/MAMELON : 1.10E-5 NcmE3/s FUITE RELEVÉE ALIMENTATION ECROU/MAMELON : 5.10E-3 NcmE3/s VIE EN LETAT</p> <p>ACCEPTÉ EN LETAT</p>	
	<p>22/5/6 : REPRISE DES ETANCHEITES DANS OPE WMS508EO. RESULTATS OK</p>	<p>A J18 (V508 - V130) - COMPLEMENT DE REMPLISSAGE WMS500EO. P FINALE 22/06/00 : 67b (PAR REGLAGE DETENDEUR) - V7 OUVERTE. LA SURVEILLANCE HEBDOMADAIRE DE LA PC VIRH AU COURS DU STOCKAGE SUR LES ELAS A MONTRE UNE STABILITE A 75b</p>	

Conclusion du clustering

Les résultats pour la “Cause” semblent probants. Les énoncés sont suffisamment réguliers pour être *clusterisés*. En revanche, le poids de la ponctuation sur cette opération est surprenant et reste à évaluer. Des méthodes de *pooling* (calcul des *embeddings* pour la représentation du texte) différentes peuvent être à envisager.

Concernant le champ “Action réalisée”, les résultats sont moins probants. On observe une plus grande diversité des énoncés, ce qui rend le *clustering* moins efficace.

Merci de votre attention

Références

- Bokinsky, H., McKenzie, A., Bayoumi, A., McCaslin, R., Patterson, A., Matthews, M. M., Schmidley, J., and Eisner, L. (2013). Application of natural language processing techniques to marine v-22 maintenance data for populating a cbm-oriented database. In AHS Airworthiness, CBM, and HUMS Specialists' Meeting, Huntsville, AL.
- Dima, A., Lukens, S., Hodkiewicz, M., Sexton, T., and Brundage, M. P. (2021). Adapting natural language processing for technical text. *Applied AI Letters*, 2(3) :e33.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1) :20–32.
- Galand, L., Kurela, M., and Clavijo, H. R. (2018). Techniques de TAL pour la recherche des "signaux faibles" et catégorisation des risques dans le REX SDF des lanceurs spatiaux. In Congrès Lambda Mu 21, " Maîtrise des risques et transformation numérique : opportunités et menaces ", Reims, France.
- Kurela, M., Bacqué, M., and Laurent, R. (2020). Classification automatique des faits techniques pour la conformité des lanceurs spatiaux. In Congrès Lambda Mu 22 " Les risques au cœur des transitions " (e-congrès) - 22e Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement, Institut pour la Maîtrise des Risques, Le Havre (e-congrès), France.
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton and Christopher D. Manning. 2020. [Stanza: A Python Natural Language Processing Toolkit for Many Human Languages](#). In Association for Computational Linguistics (ACL) System Demonstrations. 2020.
- Reimers, N. and Gurevych, I. (2019). Sentence-bert : Sentence embeddings using siamese bert-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.

```
>>> Comment est-ce que tu calcule les embeddings de phrase avec la fonction ollama
... ma.embed ?
```

La fonction ``ollama.embed`` est une API en ligne qui permet de calculer des embeddings de phrases à partir d'un texte donné. Voici comment elle fonctionne :

1. ****Tokenization**** : La fonction ``ollama.embed`` prend un texte comme entrée et le sépare en plusieurs sous-unités s'appelant "tokens" (par exemple, les mots). Cela permet de capturer les relations entre ces sous-unités.
2. ****WordPiece Tokenization**** : Si vous utilisez le mode par défaut (``ollama.embed("text", model="sentence")``), la fonction utilisera une technique appelée WordPiece Tokenization pour séparer les tokens. Celle-ci coupe les mots à partir de leur préfixe et du suffixe pour créer des sous-unités plus petites.
3. ****Embeddings calculation**** : Une fois que les tokens sont calculés, la fonction utilise un modèle d'apprentissage automatique (en l'occurrence, le modèle Sentence-Transformers) pour générer des embeddings de phrases à partir de ces tokens. Les embeddings sont des représentations numériques qui capturent les caractéristiques du texte.
4. ****Retour des résultats**** : La fonction renvoie un tableau NumPy contenant les^C

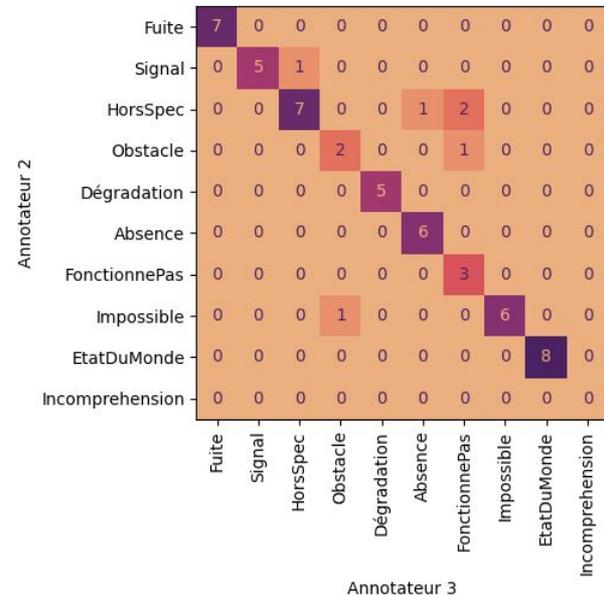
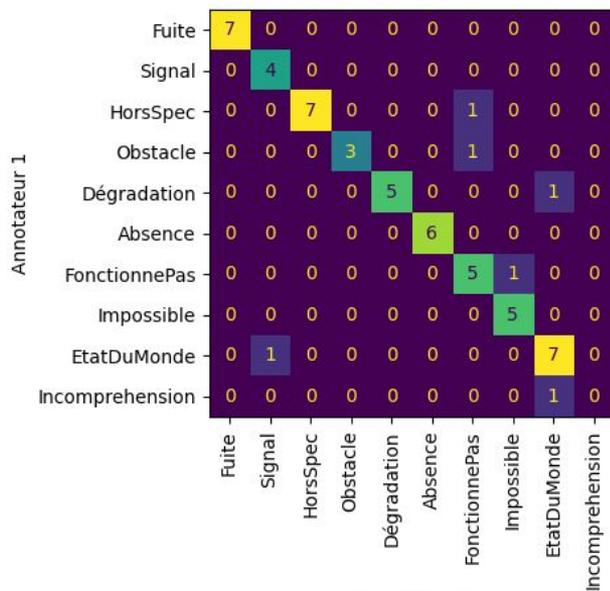
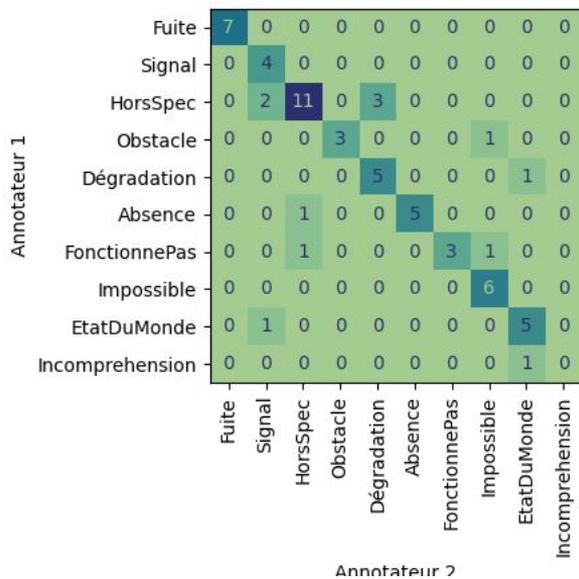
```
>>> Quel mode de "pooling" est-ce que tu utilises ?
```

Par défaut, la fonction ``ollama.embed`` utilise le mode de "pooling" appelé ****Mean Pooling****. Ce mode consiste à calculer la moyenne des embeddings des tokens pour obtenir une représentation globale du texte.

Cependant, il est possible de choisir d'autres modes de pooling en spécifiant l'option ``pool`` lors de l'appel de la fonction. Voici quelques-uns des modes de pooling disponibles :

- * ****Mean Pooling**** : La moyenne des embeddings des tokens (par défaut).
- * ****Max Pooling**** : Le maximum des embeddings des tokens.
- * ****Sum Pooling**** : La somme des embeddings des tokens.
- * ****AvgLastPooling**** : La moyenne des embeddings des tokens, mais en ignorant les tokens non significatifs (c'est-à-dire, les tokens qui ne contiennent pas d'informations utiles pour la compréhension du texte).

LA CATÉGORISATION DES SEGMENTS



	Accuracy	Kappa
A1 & A2	0,8	0,77
A1 & A3	0,84	0,81
A2 & A3	0,84	0,82

Tests - Classification automatique

Objectif : comparer l'applicabilité d'un modèle pré-entraîné sur un corpus de ce type

Classification Description – Domaine technique

Vectorisation des FT (SBERT, [Reimers et al, 2019]) (description de l'anomalie) - 1 *embedding* / FA

RL sur corpus BOW et SBERT

	Précision	Recall	F1-score
BOW	0,38	0,42	0,39
SBERT	0,29	0,40	0,26

Travaux sur l'obstacle

Difficulté : Hétérogénéité des éléments mentionnés

Approche :

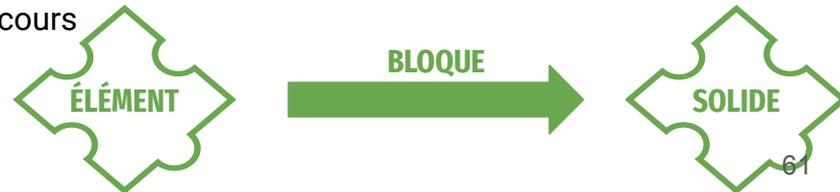
- Observation du corpus
- Élaboration du *pattern* type de l'Obstacle dans un contexte technique :
 1. **Hindrance** : Objet qui bloque
 2. **Action** : action qui n'est pas réalisée
 3. **Cible** : composant qui n'est pas modifié dû à l'obstacle

Exemple :

La porte est bloquée.	La porte est bloquée par des tapis.
Des tapis gênent.	Des tapis empêchent l'ouverture.
L'ouverture est bloquée.	L'ouverture de la porte est bloquée.
Des tapis bloquent l'ouverture de la porte.	

Méthode d'annotation envisagée : Annotation avec un LLM - tests en cours

Vépole type pour l'Obstacle :



Travaux sur l'Obstacle : Articulation avec TRIZ

Exemple

L'ACCES A LA PORTE DE LA BAIE REF-G.Y EST EGALEMENT GENE PAR DES DALLES METALLIQUES ET DES TAPIS

Étape 1

L'ACCES A LA PORTE DE LA BAIE REF-G.Y **EST EGALEMENT GENE** PAR DES DALLES METALLIQUES ET DES TAPIS → **Type : Obstacle**

Étape 2

L'ACCES A **LA PORTE DE LA BAIE REF-G.Y** EST EGALEMENT GENE **PAR DES DALLES METALLIQUES ET DES TAPIS** → **Identification des participants**

Si les éléments **Hindrance** + **Cible** sont présents dans le texte
→ obtention des éléments minimaux pour un **vépole** :

Hindrance - **BLOQUE** - **Cible**

Tests d'annotation par LLM

```
LES 3 SUPPORTS D'ECHELLE SITUES DANS LES LOCAUX TECHNIQUES LH2 ET L02 EMPECHENT LA ROTATION DES U CRYO.  
Act : rotation des U cryo  
Hin : les 3 supports d'echelle situes dans les locaux techniques Lh2 et Lo2  
Pro : la rotation des U cryo  
P2 : des U cryo
```

Méthode envisagée :

- Annotation manuelle d'un échantillon
- Annotation automatique par LLM
- Annotation automatique après redressement des données

AMBIGUITÉS

Hors spécification vs Fonctionne Pas

CAMERA ◇ N°35 ◇ FOSSE ◇ DU ◇ BIL ◇ : ◇ PROBLEME ◇ D'IMAGE.

État du Monde vs Hors Spécification

IMAGE ◇ SOMBRE ◇ CAMERA ◇ 53 ◇ DU ◇ BAF.

Obstacle vs Fonctionne Pas

ASCENSEUR ◇ HC ◇ SUPERIEUR ◇ ◇ BLOQUE ◇ AU ◇ NIVEAU ◇ +11 ◇ (TEL. ◇ INTERIEUR ◇ A ◇ L'ASCENSEUR ◇ HORS ◇ SERVICE).

Dégradation vs Hors Specification

BOULONNERIE ◇ DESSERREE ◇ AU ◇ NIVEAU ◇ DE ◇ LA ◇ RAMBARDE ◇ DE ◇ L'ESCALIER ◇ D'ACCES ◇ AU ◇ LOCAL ◇ LH2 ◇ 2.2

◇ SUD, ◇ NIVEAU ◇ 4000 ◇ DE ◇ LA ◇ TABLE ◇ 2 ◇ (VOIR ◇ PHOTOS).