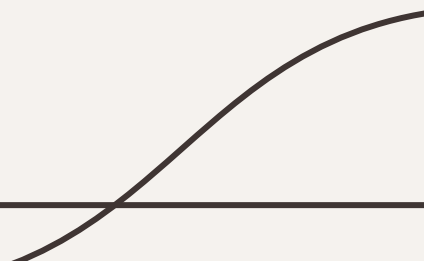




A Two-level Classification of Speech Acts for Crisis Management in Social Media

Auteurs : Enzo Laurenti , Nils Bourgon , Farah Benamara , Alda Mari ,
Veronique Moriceau , Camille Courgeon

Presenté par : Doruntina Fazliu et Iuliia Arsenteva



Plan de présentation

01

Introduction

02

Article

Présentation de l'article,
méthodes utilisés et
résultats

03

Notre recherche

Présentation de notre
travail basé sur l'article

04

Conclusion

Nos critiques et résultats



01

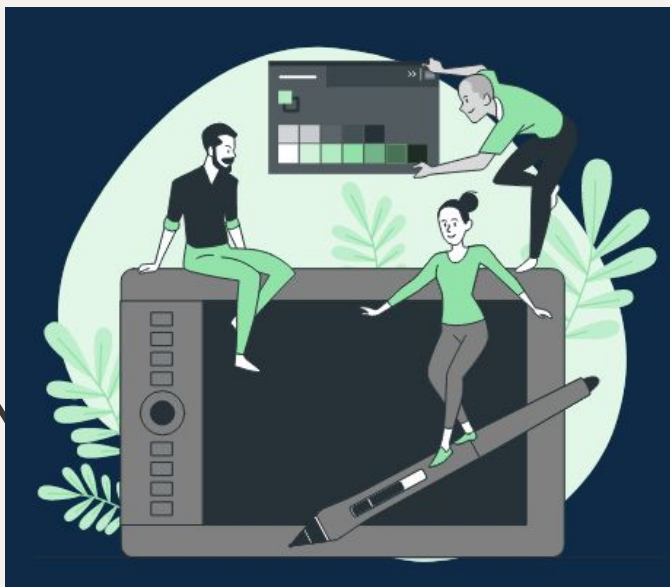
Introduction

Introduction

Nous présentons la première étude basée sur corpus pour mesurer l'impact des actes de langage dans les messages postés lors d'événements de crise dans les médias sociaux.

D'abord nous voulons:

1. mesurer l'impact de SA dans la détection d'intentions lors d'événements de crise dans des données annotées manuellement, et
 2. Explorer la faisabilité de la détection automatique de SA dans des corpus de crise.
-



02 Article

- Speech act
- Classification et annotations
- Data et Annotation
- Automatic Detection of SA

Speech Act ou Acte de Langage

Un acte de langage (ou acte de parole) est un moyen mis en œuvre par un locuteur pour agir sur son environnement par ses mots : il cherche à informer, inciter, demander, convaincre, promettre, etc. son ou ses interlocuteurs par ce moyen.



Classification et annotations

Niveau 1

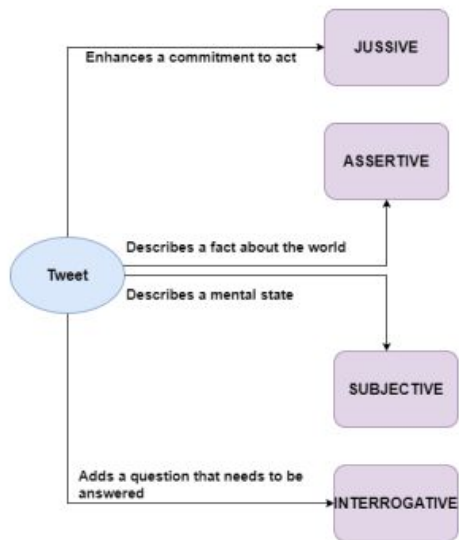


Figure 1: A classification for tweets that makes use of four illocutionary categories.

JUSSIVES - renforce l'engagement à agir
#Inondation Si vous êtes en zone inondable, découvrez comment préparer un kit de survie

ASSERTIVE - décrit un fait dans le monde
Inondations dans l'Aude : la région débloque 25 M€, le président Macron sur place lundi

SUBJECTIVE - décrit un état mental
Grosse pensée à ma Laure qui est en Martinique avec l'ouragan

INTERROGATIVE - une question dont on attend une réponse
Salut Chelsea, comment ça va, la tempête, par chez vous?

OTHERS

Simulation #3D d'une #inondation à Issy-les-Moulineaux merci a @Ubick3D pour le prêt^ #ortho3D #InterAtlas

Classification et annotations

Niveau 2

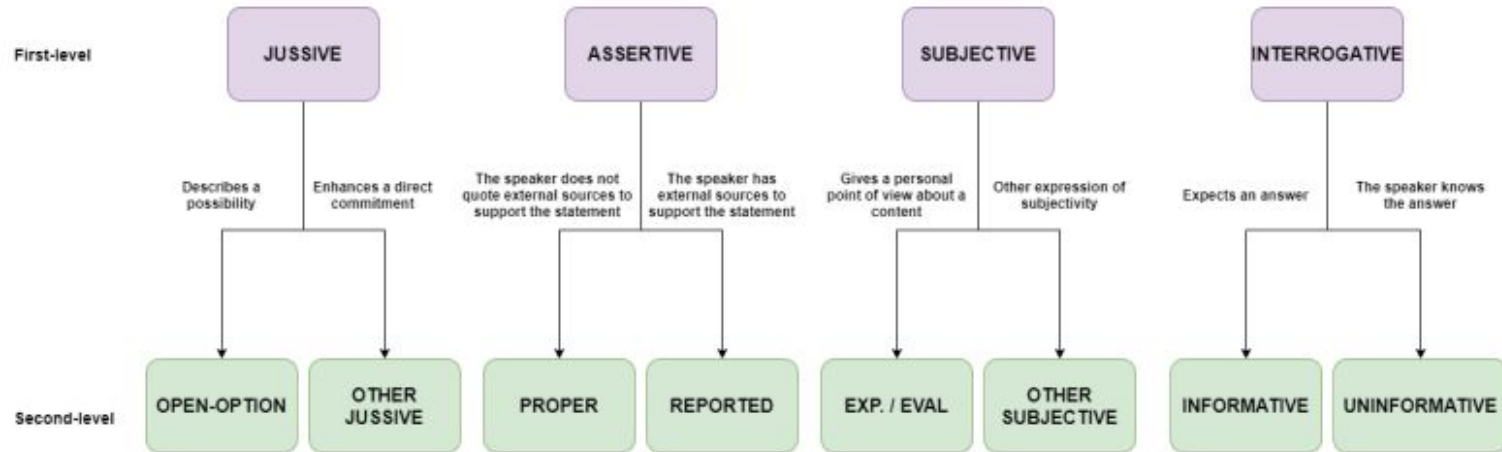
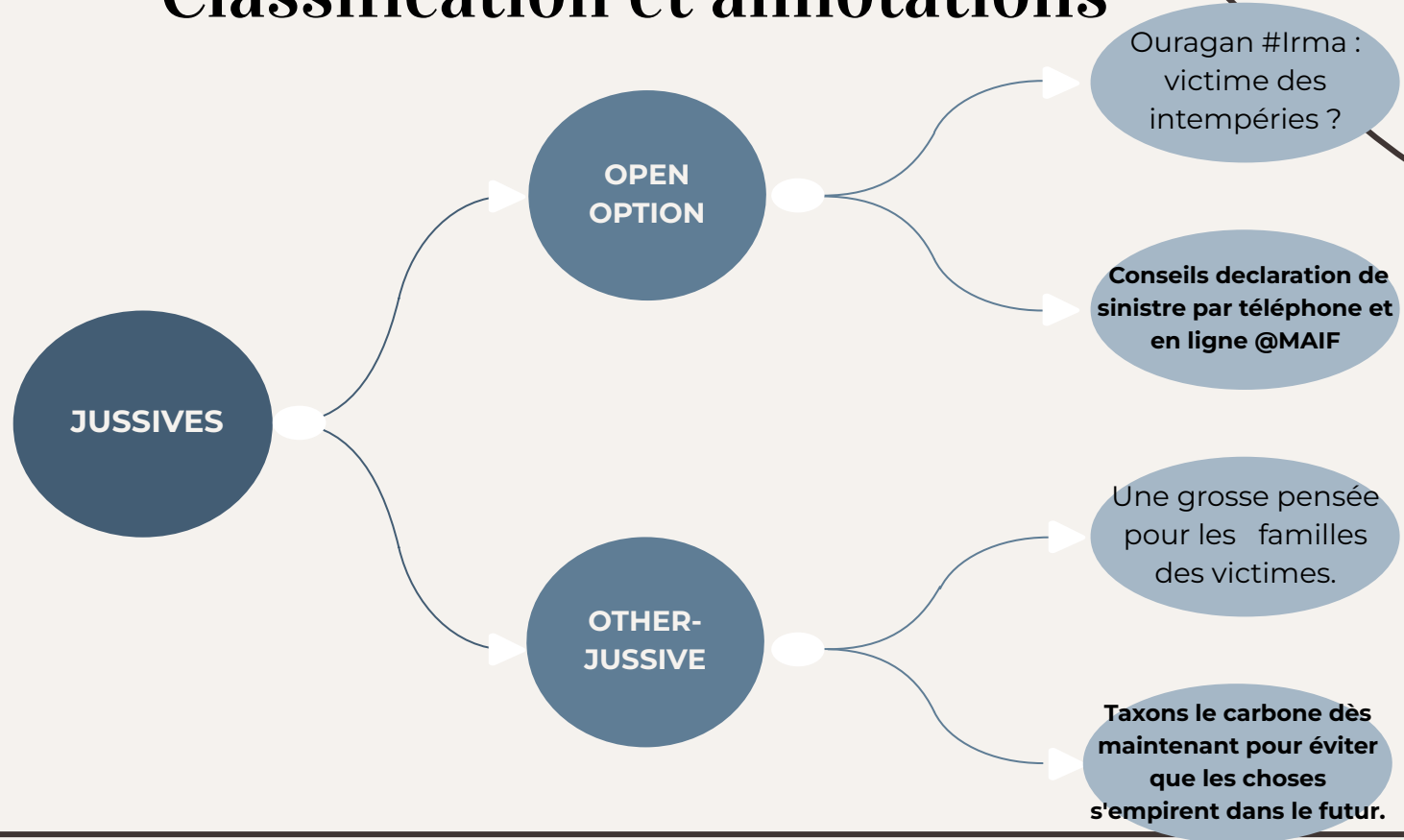
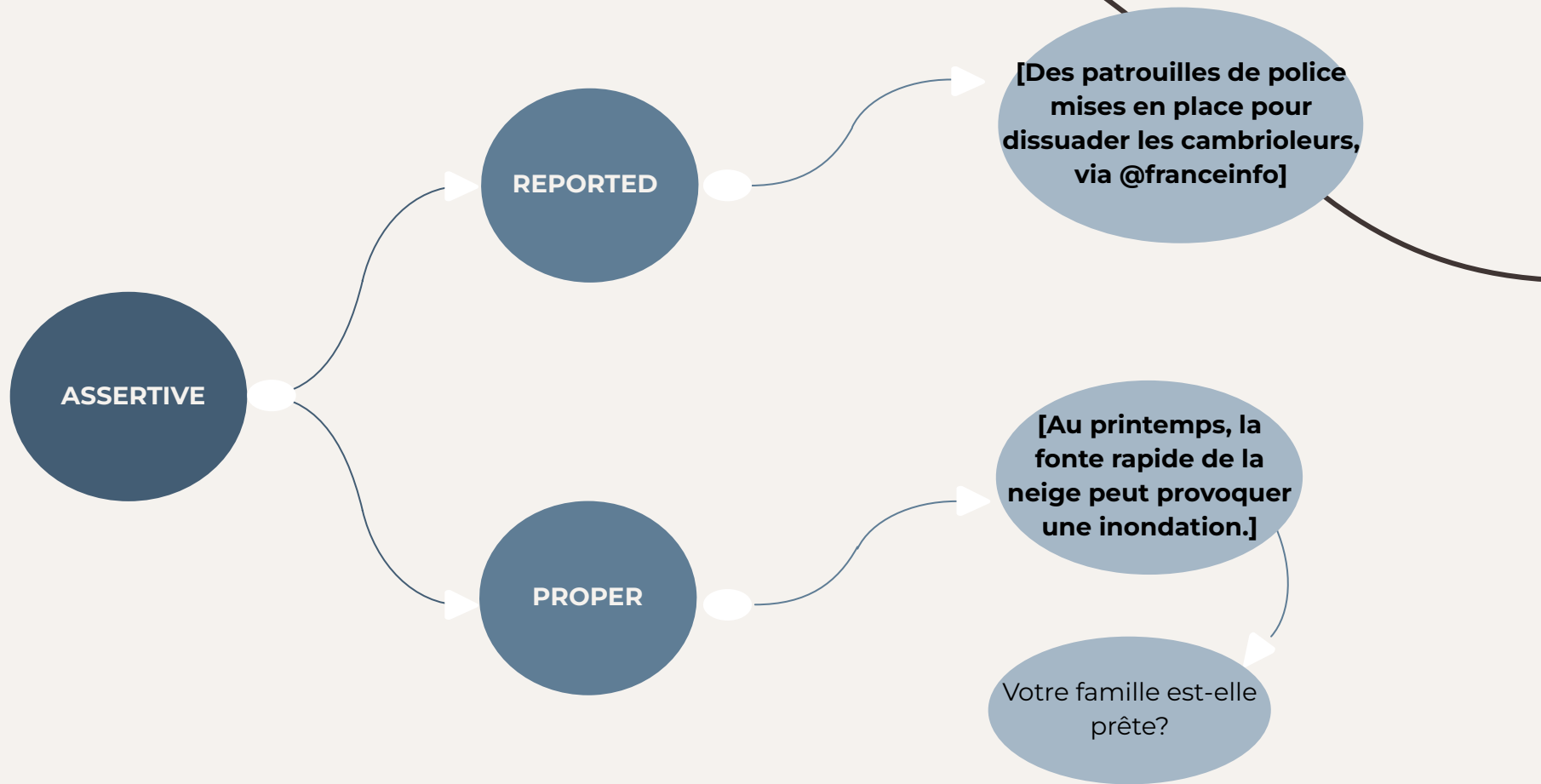
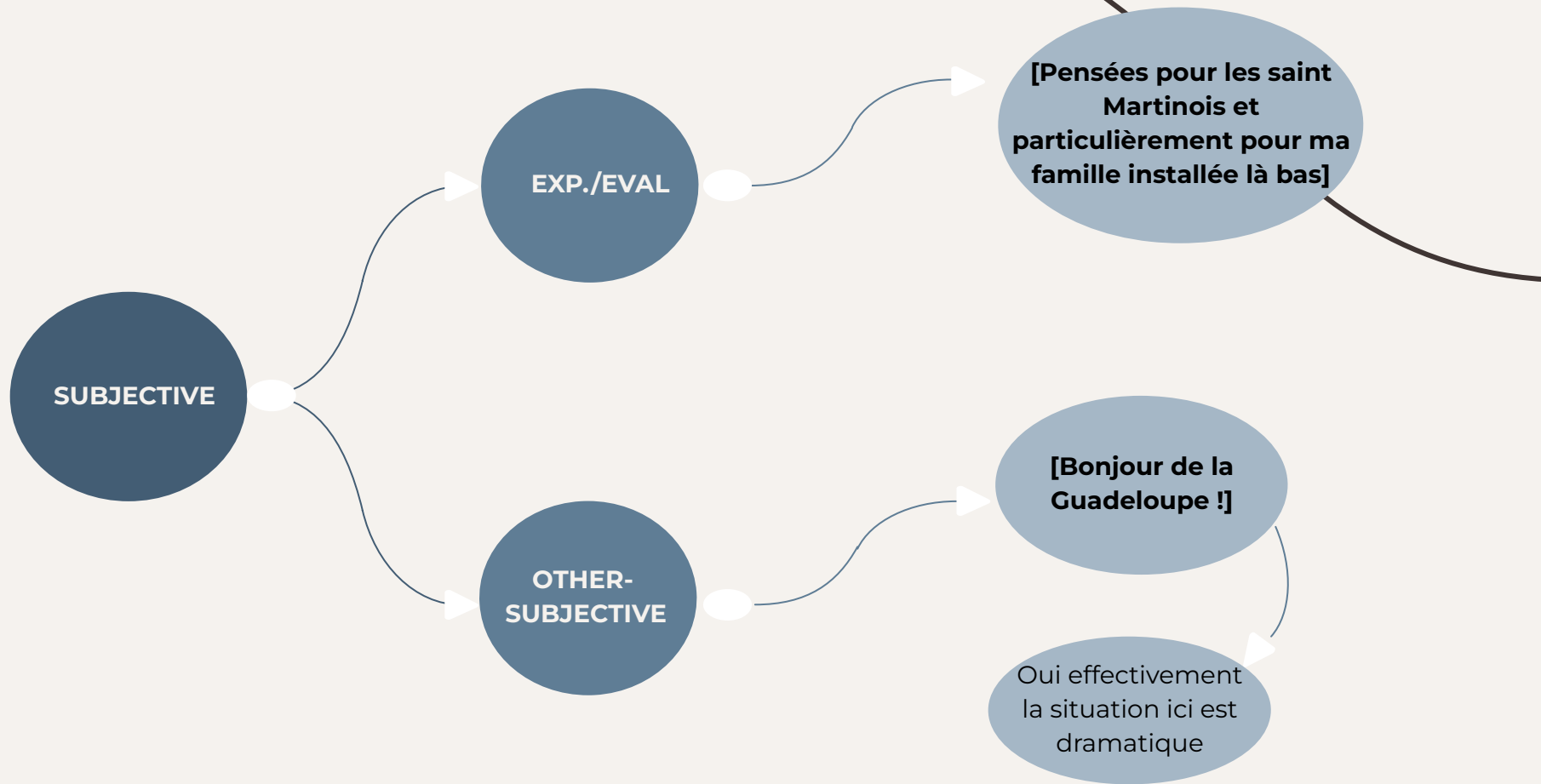


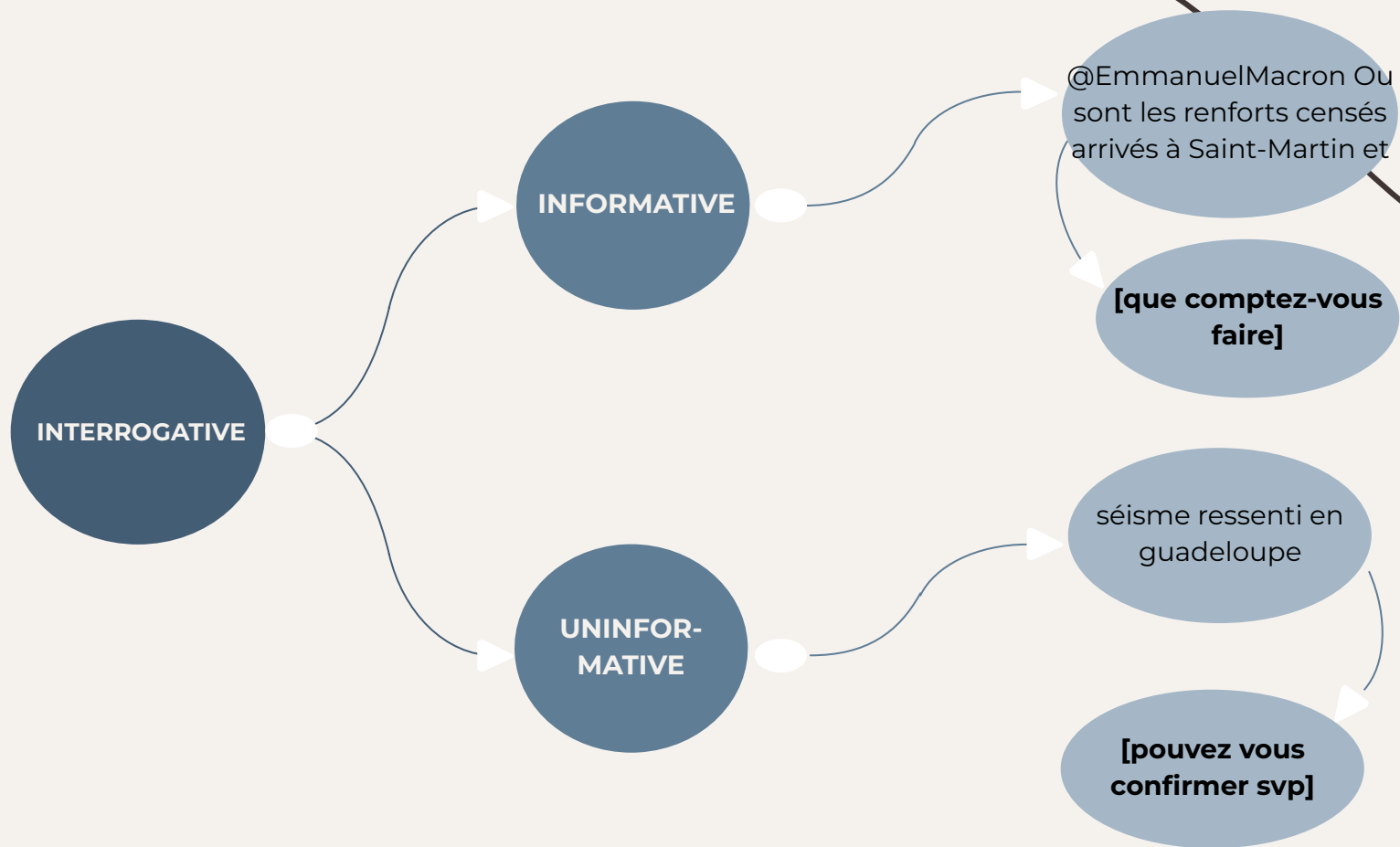
Figure 2: Two-layers annotation for tweets and inner segments.

Classification et annotations



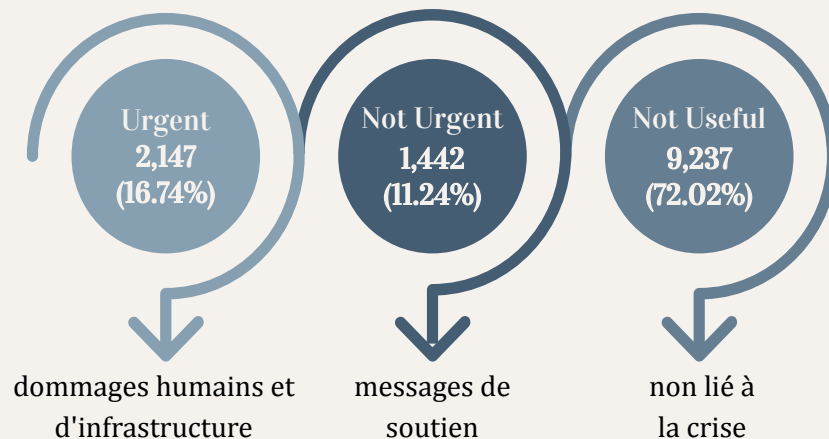






Data et Annotation

Le corpus consiste de **12 826** tweets français (Kozlowski et al., 2020) collectés à partir de mots clés dédiés aux crises écologiques survenues en France de 2016 à 2019 et postés 24h/24 avant, pendant (48h) et 72h après la crise. L'ensemble de données est fourni avec des métadonnées supplémentaires, notamment : nombre de **likes** et de **retweets**, **followers**, **suivi** de l'utilisateur.



Data et Annotation

Résultats de la campagne d'annotation

Un sous-ensemble de cet ensemble de données composé de **6 669 tweets** sont choisis et ont été annotés. Presque tous les messages **URGENTS (2 080)** et **NON URGENTS (1 401)** ont été annotés. Seulement **3 188 NON USEFUL** (afin de réduire la taille de cette classe mais de la garder majoritaire).

Cohen's Kappa=0.62

	URG	NON URG	NON USEF	TOTAL
ASSERT.	1,802	682	1,506	3,990
JUSS.	145	203	321	669
SUBJ.	106	406	976	1,488
INTERR.	20	58	145	223
OTHER	7	52	240	299
Total	2,080	1,401	3,188	6,669

Table 1: Urgency- First layer SA annotation pairs statistics.

		ASSERTIVE	SUBJECTIVE	JUSSIVE	INTERROGATIVE
Flood	Aude	718	184	84	20
	Autre	631	180	137	28
	Corse	248	73	45	23
	Total	1,597	437	266	71
Storms	Beryl	174	87	22	11
	Bruno	201	94	17	15
	Susanna	230	92	45	6
	Ulrika	170	60	43	7
	Berguitta	189	73	35	14
	Fionn Corse	238	69	28	6
	Egon	185	95	24	10
	Eleanor	208	69	26	7
	Total	1,595	639	240	76
Hurricane	Harvey	168	59	36	23
	Irma	487	251	100	36
	Total	655	310	136	59
Collapse	Marseille	143	102	27	17

Table 2: SA distribution for each crisis.

Data et Annotation

	URG	NON URG	NON USEF
JUSSIVE			
open-opt.	5.79	8.78	8.41
other.	7.85	6.96	5.31
ASSERTIVE			
report.	15.41	7.84	7.81
proper.	60.80	39.63	45.01
INTERROGATIVE			
infor.	0.22	1.66	2.42
uninfor.	1.23	3.90	4.90
SUBJECTIVE			
eval/exp.	6.89	28.36	19.14
other.	1.80	2.85	7.00

Table 3: Urgency- Second layer SA annotation pairs in percentage.

Second-level SA	%	
PROPER ASSERTION	20.69	42.47
PROPER ASSERTION + other(s) SA	21.78	
EVAL./EXPR.	8.24	21.36
EVAL./EXPR. + other(s) SA	13.12	
REPORTED ASSERTION	5.22	7.68
REPORTED ASSERTION + other(s) SA	2.46	
OPEN-OPTION	5.04	6.69
OPEN-OPTION + OTHER(S) SA	1.65	

Table 4: Distribution of most frequent second-level SA.

Détection automatique des SA

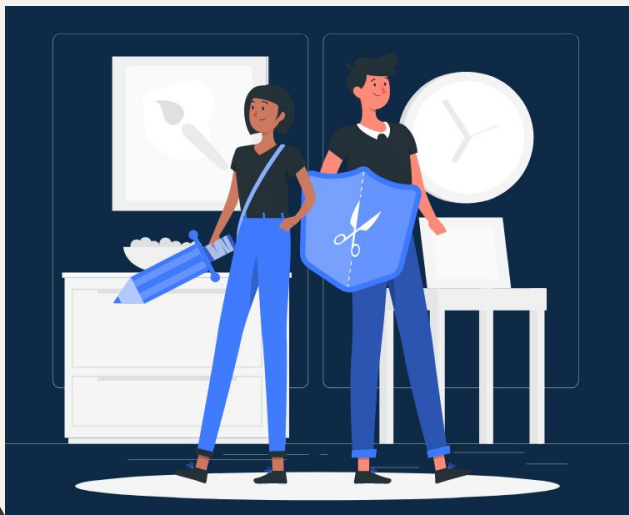
Note: Le dataset final est composé de 6,370 tweets.

Models	P	R	F
BERT _{base}	64.81	58.00	60.80
FlauBERT _{base}	72.13	66.19	68.80
CamemBERT _{base}	74.16	70.57	71.22
CamemBERT _{base} +F	75.26	70.47	72.64
CamemBERT _{focal}	75.23	71.62	72.22
CamemBERT _{focal} +F	75.66	71.95	73.55

Table 5: Overall SA classification results.

	P	R	F
ASSERT.	87.06	88.72	87.89
JUSS.	75.22	60.28	64.44
SUBJ.	72.93	77.10	66.93
INTERR.	67.44	61.70	74.96
Accuracy=81.87			

Table 6: Best model results per class.



03

Notre travail

Présentation de notre
travail basé sur l'article

Composition de dataset

3 options d'élaboration de corpus:

- Utilisation de corpus existant
- Extraction manuelle
- Extraction automatique



Extraction Automatique

```
In [8]: import tweepy
from requests_oauthlib import OAuth1Session

client = tweepy.Client(bearer_token='bearer token from dev|portal')

query = 'ouragan -is:retweet lang:fr'
tweets = client.search_recent_tweets(query=query, tweet_fields=['context_annotations', 'created_at'], max_results=100)

with open("dataset2.txt", "a") as file:
    for tweet in tweets.data:
        print(tweet.text)
        print("\n-----\n")
        print(tweet.text, file=file)
        print("\n-----\n", file=file)
        if len(tweet.context_annotations) > 0:
            #print(tweet.context_annotations)
            print(tweet.context_annotations, file=file)
```

ZES : les régies financières appelées à respecter les avantages fiscaux, parafiscaux et douaniers

<https://t.co/zL65mlbKjG>

L'ouragan Lisa, rétrogradé en tempête tropicale, est arrivé au Belize <https://t.co/2BIG3bGmpw>

Cette #Tesla (@Tesla) Model X roule en Mode Bateau en plein Ouragan Ian <https://t.co/vSRPRFACYU>

Mots Clés pour extraction

Inondation

Tremblement de terre

Tsunami

Incendie

Ouragan



Pourquoi?



La solution automatique est plus rapide



La possibilité de découvrir les bibliothèques python



Les datasets qui existent déjà ne répondent pas à nos besoins

Annotation manuelle

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	n	text	type	Iulia		doruntina				subjective	assertive	interrogative	other		
2				1st level	2nd level	1st level	2nd level							pi	sum
3		@bruno_deshayes1 @rene_chiche oui désolé si j'ai envie d'éviter une noyade lors d'inondation, satan finalement 1 c'est pas si mal	inondation	subjective	other	Subjective	other		not urgent	2	0	0	0	1	2
4		@IrreDuctible58 @rene_chiche quoi, rester dans un bâtiment lors d'inondation pour éviter une noyade ? la désobéissance généralisé ça 2 ne vous réussit pas on dirait	inondation	interrogative	uninformative	Subjective	exp/eval		not urgent	1	0	1	0	0	2
5		L'eau est entré chez une amie.. Heureusement, cette inondation a permis de découvrir qu'il y avait un serpent noir qui vivait peinard avec eux. 😭😭 Que des 3 frissons😭	inondation	subjective	expressive	Subjective	other		not urgent	2	0	0	0	1	2
6		@BernardGontran @rene_chiche on parle d'inondation, évidemment que les gens ont envie de croire qu'etre dans un bâtiment peut les 4 sauver de la noyade	inondation	assertive	proper	Assertive	proper		not urgent	0	2	0	0	1	2
7		@FPolastron @dansleslimbes @AlerteFR @VISOV1 @RatZillaS Allez prendre vos pilules pour la tension et ne venez pas pleurer quand vous serez le seul à ne pas être au courant qu'une inondation éclair 5 arrive sur votre maison.	inondation	other		Other			not urgent	0	0	0	2	1	2

kappa de Cohen = 0,82

Les résultats

	URG	NON URG	NON USEF	TOTAL
Assertive	18	1	2	21
Jussive	0	0	0	0
Subj	3	12	2	17
Interr	1	6	0	7
Other	0	5	0	5
TOTAL	22	24	4	50

Table 1: Urgency- First layer SA annotation pairs statistics

	Assertive	Subjective	Jussive	Interrogative	Other
Inondation	2	5	0	0	3
Tremblement de terre	4	4	0	1	1
Incendie	5	1	0	4	0
Tsunami	1	6	0	1	1
Ouragan	9	1	0	1	0

Table 2: SA distribution for each crisis

Détection automatique

```
df = pd.read_csv("dataset - annotation 2.csv")  
#df = pd.read_csv("ltrain.csv")  
df.head()
```

	text	jussive	assertive	subjective	interrogative	other
0	@bruno_deshayes1 @rene_chiche oui désolé si j'...	0	0	1	0	0
1	@IrreDuctible58 @rene_chiche quoi, rester dans...	0	0	1	0	0
2	L'eau est entré chez une amie.. Heureusement, ...	0	0	1	0	0
3	@BernardGontran @rene_chiche on parle d'inonda...	0	0	1	0	0
4	@FPolastron @dansleslimbes @AlerteFR @VISOV1 @...	0	0	0	0	1

Transformation de dataset dans le format pour faire l'apprentissage

Outils

$V = \frac{1}{3} \pi r^2 h$

```
class SACCommentTagger(pl.LightningModule):
    def __init__(self, n_classes: int, n_training_steps=None,
                  super().__init__())
```

$A = \pi r^2$
 $C = 2\pi r$

BERT_MODEL_NAME = 'bert-base-multilingual-cased'
tokenizer = BertTokenizer.from_pretrained(BERT_MODEL_NAME)

	30°	45°	60°
sin	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$
cos	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$
tan	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$

$\int \sin x dx = -\cos x + C$
 $\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C$
 $\int \operatorname{tg} x dx = -\ln|\cos x| + C$

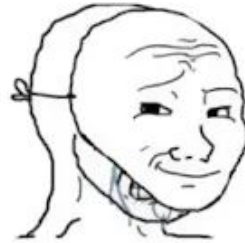
$\tan(\theta)$

$ax^2+bx+c=0$
 $a(x^2+\frac{b}{a}x+\frac{c}{a})=0$
 $x^2+2\frac{b}{2a}x+(\frac{b}{2a})^2-(\frac{b}{2a})^2+\frac{c}{a}=0$
 $(x+\frac{b}{2a})^2-\frac{b^2-4ac}{4a^2}=0$

	Name	Type	Params
0	$\frac{dx}{dt} = \frac{1}{ar(t)}$	BertModel	177 M
1	classifier	Linear	3.1 K
2	criterion	BCELoss	0

Comme la base d'apprentissage,
nous avons utilisé le script pour la
détection des commentaires
toxiques que nous avons adapté
pour la détection des actes de
langage

DESIGNERS



Look, we have similar ideas.



No! You stole my idea.

PROGRAMMERS



Man, I stole your code.



It's not my code.

Les résultats d'apprentissage

```
test_comment =  
"L'ouragan Lisa touche terre le long de la côte d'Amérique centrale \  
en tant que tempête de catégorie 1 https://t.co/QwlEcPoiVh"  
encoding = tokenizer.encode_plus(  
    test_comment,  
    add_special_tokens=True,  
    max_length=512,  
    return_token_type_ids=False,  
    padding="max_length",  
    return_attention_mask=True,  
    return_tensors='pt',  
)  
_, test_prediction = trained_model(encoding["input_ids"], encoding["attention_mask"])  
test_prediction = test_prediction.flatten().numpy()  
for label, prediction in zip(LABEL_COLUMNS, test_prediction):  
    print(f"{label}: {prediction}")
```

```
assertive: 0.47902461886405945  
subjective: 0.3917149603366852  
interrogative: 0.33085936307907104  
other: 0.2968299388885498
```

Problèmes rencontrés



Accès Twitter

	Essential	Elevated	Elevated+ (coming soon)	Academic Research
Access to full-archive search Tweets	✗	✗		✓
Access to full-archive Tweet counts	✗	✗		✓
Access to advanced filter operators	✗	✗		✓

When there's a task that can
be done manually in 10 minutes
but you find a way to automate
it in 10 days





Le bruit de dataset

Oui je suis d'accord avec vous il y'a trop de migrants en France ils faut arrêter stop stop à l'immigration même l'inondation de migrants délinquants qui ne respecte plus là France on leur doit rien vraiment rien et sa nous coûte très très cher aujourd'hui c'est stop twitter.com/Loby441/status...



Compatibilité des scripts

```
class TrainingArgumentsWithMPSSupport(TrainingArguments):  
  
    @property  
    def device(self) -> torch.device:  
        if torch.cuda.is_available():  
            return torch.device("cuda")  
        elif torch.backends.mps.is_available():  
            return torch.device("mps")  
        else:  
            return torch.device("cpu")
```

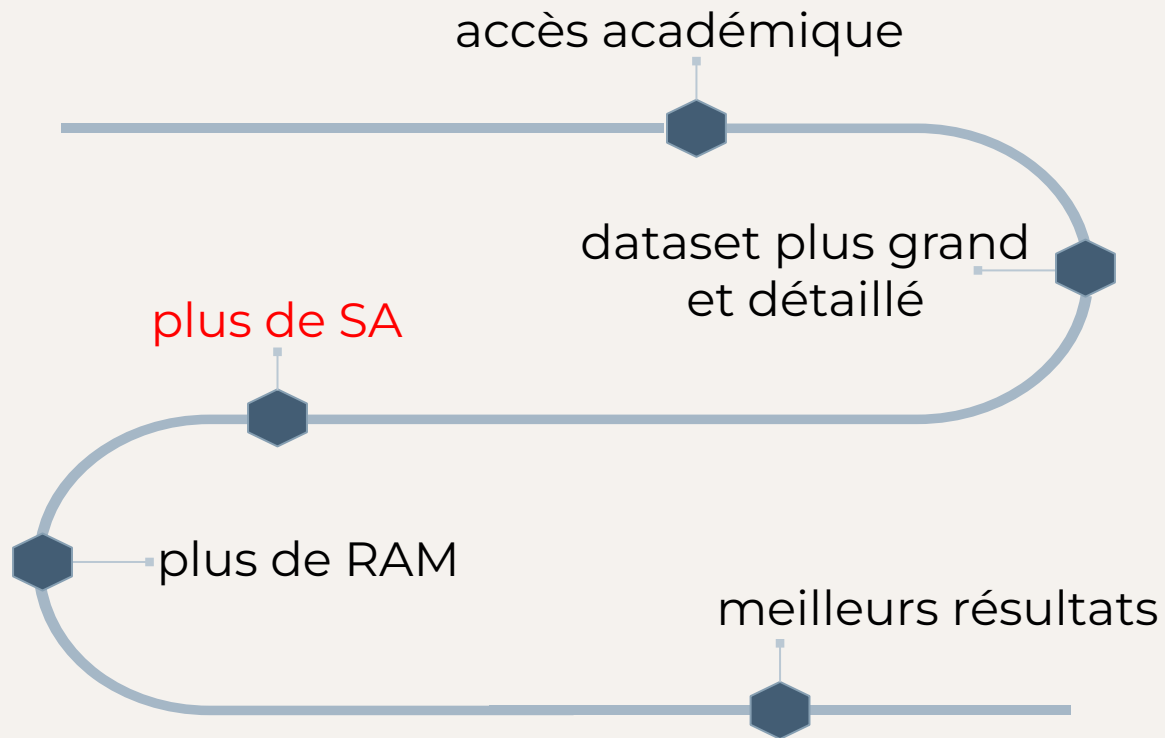
device

device(type='cpu')



04

Conclusion





**Merci de
votre
attention**

THANKS FOR YOUR ATTENTION

NO QUESTIONS PLEASE

makeameme.org