

# IA & Humour: Projet JOKER

## Interculturalité et localisation de l'humour et des jeux de mots

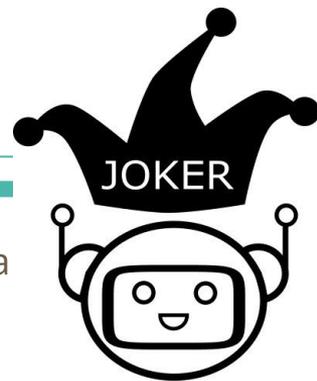
Liana Ermakova et al.

Site web: <https://motsmachines.github.io/joker/>

Email : [joker.humour.project@gmail.com](mailto:joker.humour.project@gmail.com)



\*Ce projet a bénéficié d'une aide de l'état gérée par l'Agence Nationale de la Recherche au titre du programme « Investissements d'avenir » portant la Référence ANR-19-GURE-0001





# Organizers

## France

### SHS

- Radia Hannachi
- Benoît Jeanjean
- Gaëlle Le Corre
- Élise Mathurin
- Mohamed Saki
- Adrien Couailliet
- Ludivine Grégoire
- Orlane Puchalski



### IA

- Liana Ermakova
- Anne-Gwenn BOSSER

## Étranger

- Sílvia Araújo
- Fabio Regattin
- Monika Bokiniec
- Sergio Portelli
- Ġorg Mallia
- Gordan Matas



L-Università  
ta' Malta

- Tristan Miller



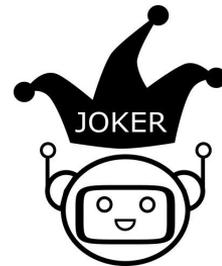
UNIVERSITÀ  
DEGLI STUDI  
DI UDINE



UNIVERSITY OF GDANSK



# Motivation



La communication interculturelle repose en grande partie sur la traduction.

L'humour reste de loin l'un de ses aspects les plus difficiles:

- références culturelles implicites
- capter le double sens

## Jeux de mots

- source d'humour courante
- romanciers, poètes et dramaturges
- titres, toponymes, anthroponymes, noms d'organisations
- publicité
- analyse de réputation
- systèmes de dialogue
- systèmes de recommandation
- détection de fake news du discours haineux
- **soustitres !!!**



# État de l'art

- La traduction est assistée par des outils technologiques
- Peu de travaux sur l'automatisation de la traduction de l'humour et des jeux de mots
- Pas de corpus parallèle
- Corpus d'humour monolingues:
  - SemEval (2017, 2020)
    - #HashtagWars: Learning a Sense of Humor
    - Detection and Interpretation of English Puns
    - Assessing Humor in Edited News Headlines
  - HaHackathon (2021): Detecting and Rating Humor and Offense
  - Autre corpus monolingues: news headlines, proverbs, British National Corpus, Yahoo!Answers etc. in English, Italian, Russian and Spanish

UBO  
Université de Bretagne Occidentale



<https://motsmachines.github.io/2021/>



JOURNÉE D'ÉTUDES

WORKSHOP

**5**  
**mars**  
**2021**

**Mots/Machines #3**  
**Les machines ont-elles**  
**le sens de l'humour ?**

**Words/Machines #3**  
**Do machines have a**  
**sense of humor?**

⌚ 8 h 30  
> 17 h 30

Mode de participation hybride pour cause de pandémie  
*Due to the COVID-19 pandemic, the conference will be held in hybrid mode*

📍 Salle C219  
FACULTÉ DES LETTRES  
ET SCIENCES HUMAINES  
20, rue Duquesne - BREST

Inscription obligatoire - Mandatory registration  
 <https://motsmachines.github.io/2021/>

[univ-brest.fr](https://univ-brest.fr)

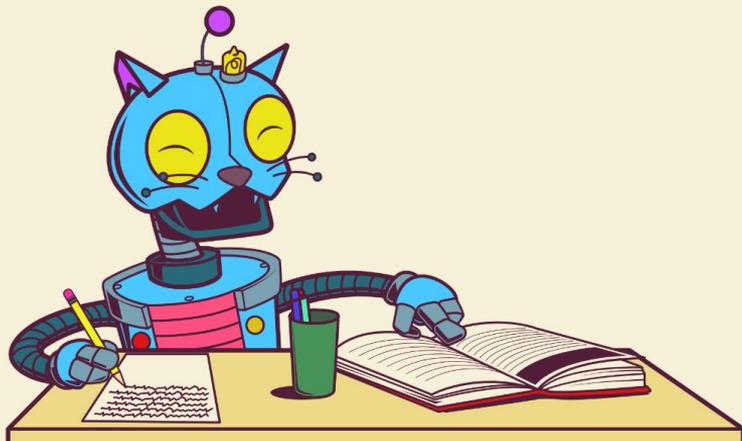
FACULTÉ  
DES LETTRES &  
SCIENCES HUMAINES

Brest  
UNIVERSITÉ

Finistère  
UNIVERSITÉ

HCTI

# PunCAT

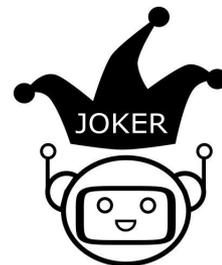


## Computational Punderstanding

Computer-Assisted Translation of Humorous Wordplay

Learn more

- <https://punderstanding.ofai.at/>
- Dr. Tristan Miller, Austrian Research Institute for Artificial Intelligence  
Vienna, Austria



# Jam Création de textes poétiques ou drôles

- <http://afia-competitions.fr/2020-Introduction/>
- Jam = programmation collaborative le soir avec pizzas
- Le Jam de 2020 est reporté à PFIA'22
  
- L'objectif serait de développer des outils qui génèrent (ou aident à générer) un texte poétique ou drôle en fonction d'une entrée qui pourra être une image/une liste de mots/une phrase...



**Afia**

Association française  
pour l'Intelligence Artificielle



# Exemple nom propre 1



DeepL Translator DeepL Pro



Translate text  
26 languages



Translate files  
.docx & .pptx

French (detected) ▾

English (US) ▾



Idéfix

Dogmatix

Alternatives:

Idéfix

Google Translate



Text



Documents

DETECT LANGUAGE

FRENCH ▾

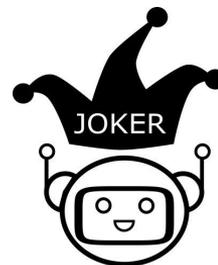


ENGLISH

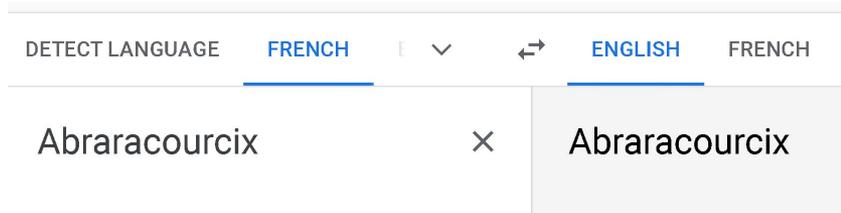
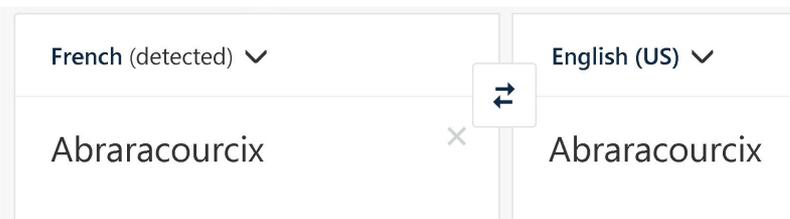
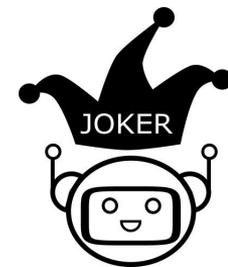
Idéfix



Dogmatix



# Exemple nom propre 2: Abraracourcix → Vitalstatistix



## Abraracourcix dans le monde [\[ modifier \]](#) [\[ modifier le code \]](#)

Comme beaucoup de personnages de la [série](#), son nom évolue selon les éditions linguistiques. .

- *Macroeconomix* en version américaine
- *Aladobix* en version finnoise
- *Vitalstatistix* en version anglophone

# Example néologisme: Ça te dirait un Symparfait?

**UNDERTALE**



ENGLISH - DETECTED

FRENCH



ENGLISH

FRENCH

RUSSIA

Would you like some  
Nice Cream?



Envie de Nice Cream ?

English (detected) ▾

French ▾

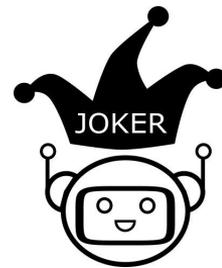


Would you like some Nice  
Cream?



Voulez-vous de la crème de  
Nice ?





# Exemple slogan publicitaire 1: Les à-côtés bien cotés

Uncle  
Ben's

English (detected) ▾

The Unexpected Side ×

English (detected) ▾

English - DETECTED FRENCH ▾ ↔ ENGLISH FRENCH R

The Unexpected Side × Le côté inattendu

French ▾

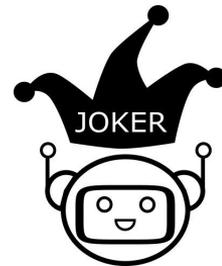
Le côté inattendu

Alternatives:

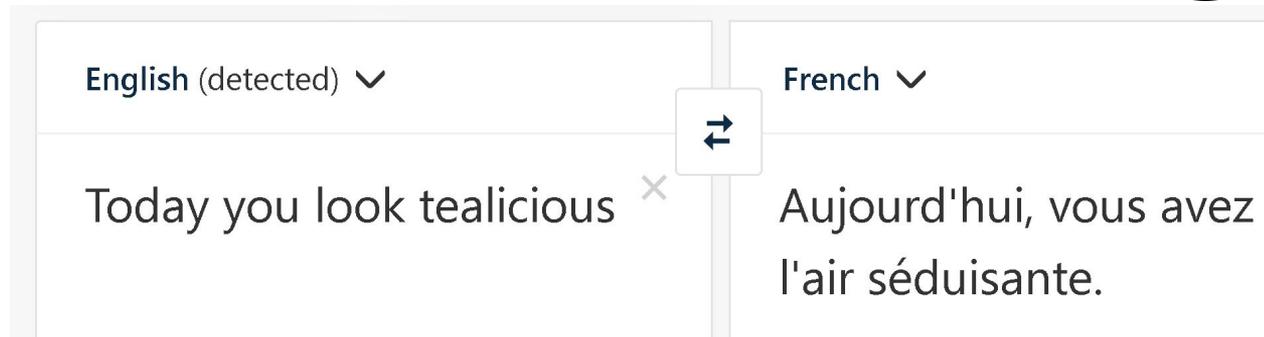
Le côté imprévu

Le côté inespéré

L'aspect inattendu



## Exemple publicité 2



ENGLISH - DETECTED

FRENCH

ENGLISH



ENGLISH

FRENCH

RUSSIAN



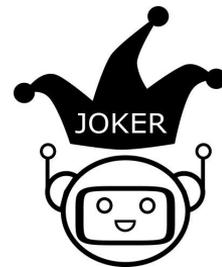
Today you look tealicious



Aujourd'hui tu as l'air délicieuse

## Exemple publicité 3

Les amoureux des chats ne boivent que du thé Mat-chat  
(Stagiaire)



ENGLISH - DETECTED

FRENCH

ENGLISH

RUSSIAN



ENGLISH

FRENCH

RUSSIAN



Cat lovers will only drink their kit-TEA|



Les amoureux des chats ne boiront que leur kit-THÉ

## Exemple littérature



C'est la raison pour quoi l'on appelle ça des **cours**, fit observer le Griffon : parce qu'ils deviennent de jour en jour plus **courts**.



ENGLISH - DETECTED

FRENCH

ENGLISH



ENGLISH

FRENCH

RUSSIAN

That's the reason they're called **lessons**,  
the Gryphon remarked, because they  
**lessen** from day to day.



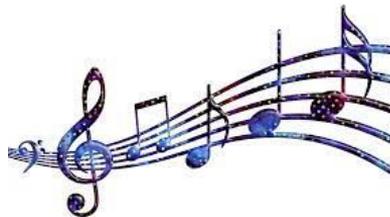
C'est la raison pour laquelle on les appelle  
leçons, remarqua le Griffon, parce qu'elles  
diminuent de jour en jour.



## Exemple tweet

- Why is music so painful?
- Because it **Hertz**

*Hertz/hurts*



- Pourquoi la musique est-elle si controversée ?
- Parce qu'elle **fait Volt-face !**

*"faire volte-face" = changement brusque et total d'opinion*

ENGLISH - DETECTED

FRENCH

ENGLISH



ENGLISH

FRENCH

RUSSIAN



Why is music so painful? Because it Hertz|



Pourquoi la musique est-elle si douloureuse ? Parce qu'il Hertz



## Exemple tweet 2

- Why did the pig leave the party early?
- Because everyone thought he was a boar!
- Pourquoi le sanglier a t-il quitté la fête ?
- Parce que tout le monde le trouvait laie



ENGLISH - DETECTED

FRENCH

ENGLISH

RUSSIAN



ENGLISH

FRENCH

RUSSIAN



Why did the pig leave the party early?  
Because everyone thought he was a boar!



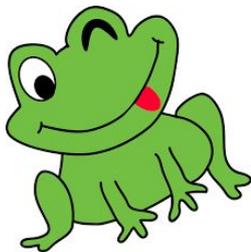
Pourquoi le cochon a-t-il quitté la fête plus tôt ? ☆  
Parce que tout le monde pensait que c'était un  
sanglier !



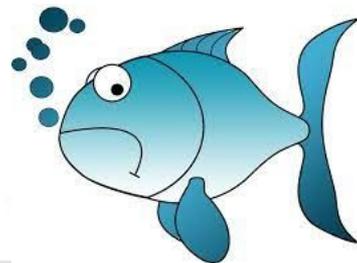
## Exemple tweet 3

- What happened when the frog's car broke down?
- First he tried to jump, then got **toad.**

*toad/towed (to pull a car)*



- Que fait un saumon pour doubler une truite ?
- **Une queue de poisson.**



ENGLISH - DETECTED

FRENCH

ENGLISH



ENGLISH

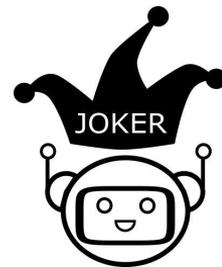
FRENCH

What happened when the frog's car broke down?  
First he tried to jump, then got toad.

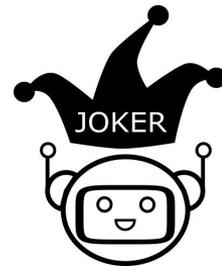


Que s'est-il passé lorsque la voiture de la grenouille est tombée en panne ?  
Il a d'abord essayé de sauter, puis s'est retrouvé avec un crapaud.

# Application de l'IA = données + fonction de coût



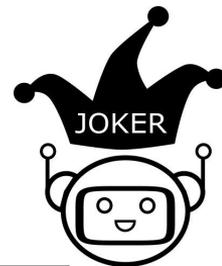
- Jeux de données
  - corpus parallèle
- Métriques d'évaluation
  - BLEU (bilingual evaluation understudy) = taux du chevauchement vocabulaire entre la traduction candidate et la traduction de référence
    - n'est applicable pour évaluer la traduction du langage créatif
  - réflexion sur les mesures d'évaluation du langage créatif potentiellement automatisables
    - est-ce que le jeu de mots est présent dans la traduction?
    - est-ce que le champ sémantique est préservé dans la traduction?
- Proposition d'un atelier CLEF + shared tasks
  - Conference and Labs of the Evaluation Forum - Information Access Evaluation meets Multilinguality, Multimodality, and Visualization
  - <https://clef2022.clef-initiative.eu/index.php?page=Pages/cfl.html>



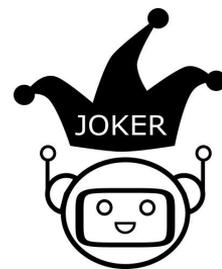
# Shared tasks

- Pilot task 1:
  - Classify and explain a given punning construction in a proper noun or a neologism
  - Evaluation metrics:
    - Classification - accuracy
    - Explanation - exact match (against the gold-standard)
- Pilot task 2:
  - Translate a given punning construction in a proper noun or a neologism from English into French
- Pilot task 3:
  - Translate a given punning phrase from English into French

# Planning

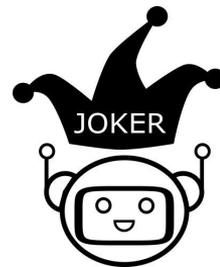


Summer 2021	CLEF workshop proposal
November 2021	Participant registration opens
February - beginning of March	Dataset release
End of April 2022	Participant registration closes
Beginning of May 2022	Evaluation Cycle
End of May 2022	Submission of Participant Papers
September 2022	CLEF conference



## List of Topics (not exhaustive)

- Humour translation
- Human-computer interaction in pun/humour translation
- Humour/pun analysis in social networks
- Humour classification
- Humour/pun rating
- Humour detection
- Evaluation metrics
- Pun translation
- Pun detection
- Joke generation
- Pun generation
- Translation error types and resultant humour
- Limits of machine translation and CAT systems in humour/pun translation
- ...



**Merci pour votre attention!**  
**Questions? Suggestions?**  
**Participez à nos tâches pilotes JOKER\* !**

Site web: <https://motsmachines.github.io/joker/>

Email : [joker.humour.project@gmail.com](mailto:joker.humour.project@gmail.com)

\*Ce projet a bénéficié d'une aide de l'état gérée par l'Agence Nationale de la Recherche au titre du programme « Investissements d'avenir » portant la Référence ANR-19-GURE-0001