Overview of the JOKER* Workshop: Automatic Wordplay and Humour Translation

Liana Ermakova et al.

Site web: **http://www.joker-project.com/**
Email : **contact@joker-project.com**

Organizers

- Liana Ermakova
- Anne-Gwenn Bosser
- Tristan Miller
- Fabio Regattin
- Sílvia Araújo
- Claudine Borg
- Radia Hannachi
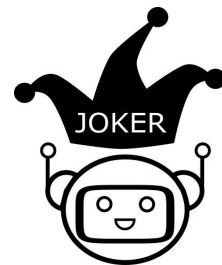- Benoît Jeanjean
- Gaëlle Le Corre
- Élise Mathurin

Students

- Adrien Couaillet
- Ludivine Grégoire
- Orlane Puchalski
- Paul Campen
- Aurianne Damoy
- Albin Digue
- Julien Boccou
- Loic Glémarec

Motivation

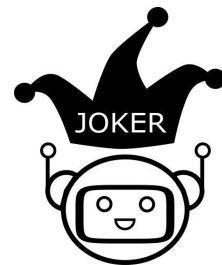Humor remains one of the most difficult aspects of intercultural communication & translation:

- cultural references
- creativity, syntactic and lexicographic features of languages

Puns

- a common source of humor in jokes and comedy shows
- novels, poems and plays
- social listening & reputation monitoring
- dialogue systems
- recommendation systems
- fake news and hate speech detection
- used for their attention-grabbing or mnemonic value in titles, place names, organizations, advertising, etc.

MT & CAT systems do not have any specific support for humour

CLEF 2022 BOLOGNA — JOKER
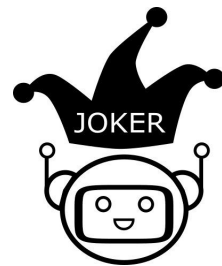
ENGLISH - DETECTED | FRENCH | ENGLISH ⌄ ⇄ ENGLISH | **FRENCH** | RUSSIAN ⌄

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

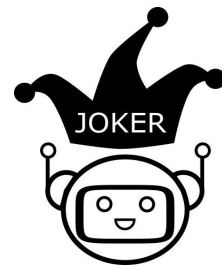C'est la raison pour laquelle on les appelle leçons, remarqua le Griffon, parce qu'elles diminuent de jour en jour.

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.

Technical & evaluation challenges

- Assumption: A word has single meaning in a sentence, but wordplay exploits the confrontation of similar forms but different meanings

- AI requires a quality and quantity of training data

- Humoristic neologism are new words!

- MT evaluation metrics measure the proportion of shared words

- No parallel data for pun translation exists

Parents: If all your friends jumped into the well, Will you ?
Kid: No !

Machine learning Algorithm: YES !

Parallel dataset
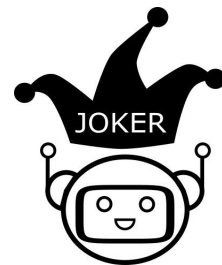
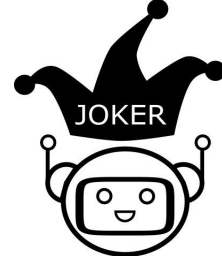- Manually collected existing translations (>2,000): video games, literature, advertising, … ✔
- Our translations:
  - Puns from the Internet (master students in translation interns) ✔
  - Translation contest ✔
    - English -> French
    - SemEval-2017 Task 7: Detection and Interpretation of English Puns (Miller et al., 2017)
  - Automatic detection, localization & translation alignment of puns from literature ✘

Pilot tasks

- **Pilot task 1:**

  - Classify and explain a given wordplay

- **Pilot task 2:**

  - Translate a given wordplay in a proper noun or a neologism from English into French

- **Pilot task 3:**

  - Translate a given wordplay phrase from English into French

- **Unshared task:**

  - We welcome any submission that uses our data!

Statistics

- 49 registered teams

Team	Task 1	Task 2	Task 3	Total runs
FAST_MT	2	1	1	4
Cecilia	1	1	2	4
Agnieszka	2	1	0	3
eBIHAR	4	0	0	4
TEAM_JOKER	0	1	0	1
LJGG	0	0	2	2
Humorless	0	0	1	1
Total runs	9	4	6	19

CLEF TASK 1
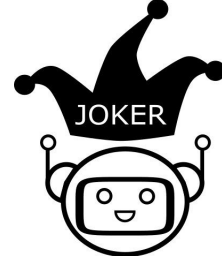
- Classify wordplay
  - HORIZONTAL/VERTICAL: co-presence of source and target of the wordplay.
    - Horizontal wordplay ex.: "They're called **lessons** because they **lessen** from day to day"
    - Vertical wordplay ex.: "How do you make a **cat drink**? Easy: put it in a liquidizer"
  - MANIPULATION_TYPE: Identity/Similarity/Permutation/Abbreviation
  - MANIPULATION_LEVEL: Sound/Writing/Other
  - CULTURAL_REFERENCE: (True/False) One has to be aware of some extra-linguistic factors to understand some instances of wordplay
- Provide lexical-semantic interpretations
- EN
- **Train: 531/ Test: 4,517**
- 4 teams
- 7 runs

Participants' results for TASK 1
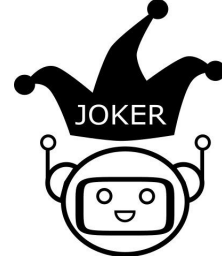
	LOCATION MANIP.	TYPE MANIP.	LEVEL
FAST_MT		1035	2437
FAST_MT_updated	1455	1667	2437
Cecilia_task_1_run5	1484	1541	2437
Agnieszka_task1_t5	1554		
eBIHAR_en		1392	2437
eBIHAR_en_tfidf_wp		1083	2437
eBIHAR_en_tfidf_wp_preprocessed		536	2437

TASK 1 lessons learnt
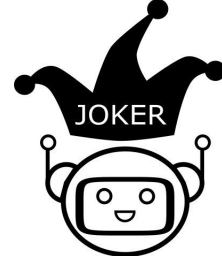
- Separate phrase-based and term-based wordplay
- Separate datasets in different languages

Phrase-based

- wordplay <u>location</u> is important
  - usually this is the <u>last</u> word
- wordplay <u>interpretation</u> is important
- remove offensive wordplay
- for unbalanced classes:
  - either remove rare classes
  - or add other types of wordplay
  - <u>or do not classify wordplay at all</u>

Term-based

- wordplay <u>interpretation</u> is important
- No need to locate wordplay
- Classification into CONVENTIONAL FORM and OFFENSIVE is not applicable

CLEF TASK 2
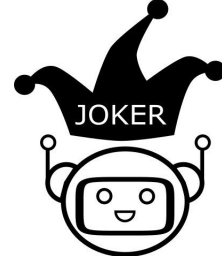
- Translate single words containing wordplay
- EN → FR
- Train: 1,161
- Test: 284

id	en	fr
noun_12	dungbomb	bombabouse
noun_13	Seel	Otaria
noun_522	Dragunfire	Feu du Dragun

POKÉPÉDIA

TASK 2 results

	FAST_MT	TEAM_JOKER	Cecilia	Agnieszka
total	284	284	284	242
not translated	0	0	0	0
official	250	159	216	230
non-official	34	125	68	12
lexical field preservation	16	13	5	0
sense preservation	13	11	5	0
comprehensible terms	26	59	16	2
wordplay form	3	12	3	1

TASK 2: Machine translations different from official ones

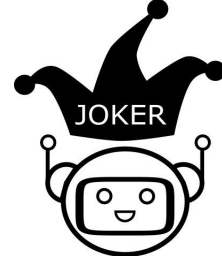| EN	FR	French Version
Orbeetle	Orbétain	Astronelle
Ribombee	Ribombe	Rubombelle
Crabrawler	Crabaque	Crabagarre
Celesteela	Cesteela	Bamboiselle
Celesteela	Célésteela	Bamboiselle
Appletun	lio	Dratatin
Appletun	loquet	Dratatin
Corviknight	Corvaillus	Corvallus
Primarina	Primarin	Oratoria
Oompa-Loompas	Oompas	Oompa-Loompas

| EN	FR	French Version
Dolores Umbridge	Dolores Ombrage	Ombrage
Black Hole Gun	Gun à l'épaule noire	Arme à trou noir
Black Hole Gun	Pistolet trou noir	Arme à trou noir
Weasleys' Wildfire Whiz-bangs	Feux de Weasleys ufs feux	Feuxfous Fuseboum
Weasleys' Wildfire Whiz-bangs	Wildfire Whiz-bangs	Feuxfous Fuseboum
Weasleys' Wildfire Whiz-bangs	'Bad-bangs	Feuxfous Fuseboum
Lickable Wallpaper Walls	Lickable Wallpaper Wall	Papier peint qui se lèche
Lickable Wallpaper Walls	Murs de papier peint léchables	Papier peint qui se lèche
Lickable Wallpaper Walls	Walls de papier à papier	Papier peint qui se lèche
Ragnar the Rock	Ragnar le rocher	Ragnar le Roc

CLEF TASK 3

- Translate entire phrases that subsume or contain wordplay
- EN → FR
- Train: 1,239 (EN) - 5,115 (FR)
- Test: 2,378
- ⇐ 500 EN puns translated into FR (~5,000 translations) ⇐ gamification

- 7 runs
- 5 teams
- 11671 different translations
- str.match('.+[?.!"]\s*$')

TASK 3 participants' results

|	LJGG DeepL	FAST_MT	LJGG auto	Cecilia run 1	Humorless	Cecilia run 3
total	2378	2378	2378	2378	2378	2378
valid	2324	2120	2264	2343	384	7
not translated	39	103	206	49	22	2
nonsense	59	220	349	51	297	3
syntax problem	17	58	46	41	6	0
lexical problem	25	79	78	52	10	0
lexical field preservation	2184	1739	1595	2155	118	6
sense preservation	1938	1453	1327	1803	100	6
comprehensible terms	1188	867	827	744	56	5
wordplay form	373	345	261	251	19	1
identifiable wordplay	342	318	240	243	16	1
over-translation	3	1	9	13	0	0
style shift	9	12	4	4	0	0
humorousness shift	930	765	838	1427	68	4

TASK 3 result statistics

•
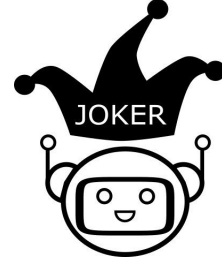
All	Untranslated	Nonsense	Truncated	Lexical field preservation	Meaning preservation	Wordplay presence	Expression/term understandable	Identifiable wordplay	Over-translation	Register shift	Funniness level shift	Syntax problem	Lexicon problem
6850	352	910	327	5173	4429	892	2768	817	13	22	2528	113	166
	5%	13%	5%	76%	65%	13%	40%	12%	0%	0%	37%	2%	2%

13% of automatically translated wordplay are successful

- **Mainly literal translation**

- A train load of paint derailed. Nearby businesses were put in the red.
- Un train de peinture a déraillé. Les entreprises voisines ont été mises dans le rouge.

- An undertaker can be one of your best friends, he is always the last one to let you down.
- Un entrepreneur peut être l'un de vos meilleurs amis, il est toujours le dernier à vous laisser tomber.

- Before he was hired as a short order cook they grilled him.
- Avant qu'il ne soit embauché comme cuisinier à commande courte, ils l'ont grillé.

- **90% of human translated wordplay are successful**

Surprising success & Anglicisms
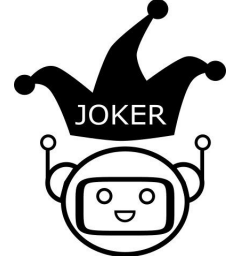
- Success comes in cans, failure comes in cant's
- Le succès c'est dans les canons, le pétrin c'est dans les canettes.

- Wal - Mart Is Not the Only Saving Place. Come On In.
- Le clerc n'est pas le seul à faire des économies.

- Gunpowder was invented in a flash.
- La poudrière a été inventée dans un flash.

- When the beekeeper moved into town he created quite a buzz.
- Lorsque l'apiculteur s'est installé en ville, il a créé un véritable buzz.

Future work
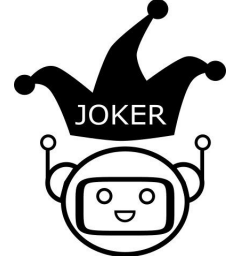
- Task 2 will be dropped out ?

- A new task: Pun detection in English and French ?

- Spanish?

- Automatic data collection?

- Improvement of automatic interpretation evaluation?

Program - Wednesday 7th September (ROOM F)

- 13:30 - 13:40 Welcome talk
- 13:40 - 14:40 Invited talk by **Paolo Rosso**

"Humour, Irony and Sarcasm: Often Lost in Translation"

- 14:40 - 18:00 participants' presentations
- 18:00 - 18:50 Round table and JOKER 2023 discussion
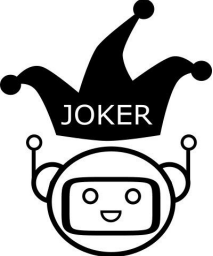
Thank you!

Participate in our tasks !

Website: http://www.joker-project.com/

Email : contact@joker-project.com

Twitter: https://twitter.com/joker_research

Google Group: https://groups.google.com/u/4/g/joker-project