JOKER CLEF Workshop: Automatic Pun and Humour Translation

Liana Ermakova¹, Tristan Miller², Orlane Puchalski¹, Fabio Regattin¹, Élise Mathurin¹, Silvia Araújo⁴, Anne-Gwenn Bosser⁵, Claudine Borg⁶, Monika Bokiniec⁷, Gaelle Le Corre⁸, Benoît Jeanjean¹, Radia Hannachi⁹, Gorg Mallia¹⁰, Gordan Matas¹⁰, and Mohamed Saki¹

1 Université de Bretagne Occidentale, HCTI EA-4249/MSHB, 29200 Brest, France
2 Austrian Research Institute for Artificial Intelligence, Vienna, Austria
3 Dipartimento DILL, Università degli Studi di Udine, 33100 Udine, Italy
4 Universidade do Minho, CEDER, Rua da Universidade, 4710-057 Braga, Portugal
5 École Nationale d’Ingénieurs de Brest, Lab-STICC CNRS UMR 6285, France
6 University of Malta, Madia MSD 2020, Malta
7 University of Gdańsk, Poland
8 Universidade de Brege Occidentale, CRBIC, 29200 Brest, France
9 Université de Bretagne Sud, HCTI EA-4249, 56321 Lorient, France
10 University of Split, Croatia

OBJECTIVE: Translators & computer scientists ⇒ evaluation framework = data + metrics development ⇒ foster work on automation of wordplay translation

MOTIVATION:
Humor is 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 and is widely used for their attention-grabbing or mnemonic value

CHALLENGES:
• Machine Translation & Computer Assisted Translation systems do not have any specific support for humour
• Assumption: A word has single meaning in a sentence. Wordplay ≠ 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

PILOT TASK 1
To classify wordplay
- HORIZONTAL/VERTICAL: 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: HORIZONTAL/VERTICAL
- CULTURAL_REFERENCE: True/False
One has to be aware of some extra-linguistic factors to understand some instances of wordplay.

To provide lexical-semantic interpretations
- EN
- Train: 531
- Test: 4,517

PILOT TASK 2
To translate single words containing wordplay
- EN → FR
- Train: 1,161
- Test: 284

PILOT TASK 3
To 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 translation) ⇒ gamification

OPEN TASK
We welcome any other submissions that use our data!