

# Overview of the JOKER Lab: Automatic Wordplay Analysis

Liana Ermakova   Tristan Miller   Anne-Gwenn Bosser  
Victor Preciado   Grigori Sidorov   Adam Jatowt



CLEF 2023, Thessaloniki, Greece, September 20, 2023

# JOKER Track Motivation



- Humor remains one of the most difficult aspects of intercultural communication & translation
- Wordplay is a common form of humor that can crop up in almost any type of discourse
- Perennial topic of scholarship in literary criticism, language education, and translation studies → need for specialized text retrieval and classification technology
- Important for opinion mining, conversational agents etc.
- Usual assumption: A word has a single meaning in a sentence, but wordplay exploits the confrontation of similar forms but different meanings
- AI requires quality and quantity of data for training and testing
- MT evaluation metrics measure the proportion of shared words

# Top “easiest” punning words



---

## EN

Martians welcome. We have **space** for everyone.

A lot of trees were dying, but they needed to figure out the **root** of the problem.

She was suspected of stealing a brooch but they couldn't **pin** it on her.

Well drilling is a **deep** subject.

The inept mathematician couldn't **count** on his friends.

---

## FR

Bienvenue les extraterrestres ! Installez vous, on a créé ces **espaces** détente pour vous.

De nombreux arbres mouraient mais personne ne trouvait la **racine** du mal qui les rongait.

Elle s'est fait **épingler** pour une histoire de broche volée.

Le forage de puits est un sujet **profond**.

Un mathématicien qui ne peut **compter** sur ses amis n'est pas un mathématicien...

---

# Goals



- To provide appropriate reusable **data** and **benchmarks** for automatic wordplay analysis.
- To provide a discussion platform to address **technical & evaluation** challenges of automatic wordplay analysis
- **Use cases:**
  - Computer-Assisted Translation of wordplay
  - Corpus-based analysis of wordplay in the humanities
    - literary criticism
    - language education
    - translation studies
    - humor studies
  - Wordplay-aware Information Retrieval

# JOKER@CLEF Shared Tasks



- Broken down into three tasks:
- Task 1: Detection of wordplay in English, French, and Spanish;
- Task 2: Location and interpretation of wordplay in English and French;
- Task 3: Translation of puns from English to French and Spanish.
- We also welcome runs within the unshared task with other, possibly novel, use cases, such as pun generation or humorousness evaluation.

# Statistics on submitted runs by task



Team	Task 1: Detection			Task 2.1: Location			Task 2.2: Interpret.		Task 3: Translation		Total
	EN	FR	ES	EN	FR	ES	EN	FR	EN-FR	EN-ES	
<i>Croland</i>	1	1	1	1	1	1	1		1	1	9
<i>LJGG</i>	3	3	3						4	5	18
<i>Les_miserables</i>	3	3	3	3	3	3	1				19
<i>MiCroGerk</i>	6			6			4			7	23
<i>Smroltra</i>	7	7	7	4	4	4	6		6	6	51
<i>TeamCAU</i>	6			3					3		12
<i>TheLangVerse</i>	1								1	1	3
<i>ThePunDetectives</i>	6			5					2	2	15
<i>UBO</i>	1	1	1	1	1	1	1		3	3	13
<i>UBO-RT</i>							1	1			2
<i>AKRaNLU</i>	2	2	2	2	2	2	1	1			14
<i>Innsbruck</i>	3										3
<i>NPalma</i>	1		1							2	4
<b>Total</b>	40	17	18	25	11	11	15	2	20	27	186

# Task 1: Wordplay Detection



- *Wordplay detection* is a binary classification task where the goal is to distinguish between texts containing an instance of wordplay and texts without it.
- (1) When the church bought gas for their annual barbecue, proceeds went from the sacred to the **propane**. - *propane/profane*
  - (2) I used to be a banker but I lost **interest**. - *interest*

# Task 1: Data & Evaluation



- Data
  - English (\*)
    - Train: SemEval 2017, 2021
    - Test: Manually augmented data
  - French (\*)
    - Manually and automatically translated puns from English
    - Manually augmented data
  - Spanish
    - Manually and automatically translated puns from English
    - Manually augmented data
- Evaluation measures:
  - Pun detection: P, R, F1, Acc
  - Pun location: Acc

---

\* L.Ermakova, A.-G.Bosser, A.Jatowt, & T.Miller. *The JOKER Corpus: English-French Parallel Data for Multilingual Wordplay Recognition. SIGIR '23*



# Task 1 data statistics



Language	Train			Test		
	Positive	Negative	Total	Positive	Negative	Total
English	3,085	2,207	5,292	809	2,374	3,183
French	1,998	2,001	3,999	5,308	7,565	12,873
Spanish	855	1,139	1,994	952	1,289	2,241

- Corpora are homogeneous in terms of the **length** of wordplay and non-wordplay instances
- Corpora are homogeneous in terms of **vocabulary** across wordplay and non-wordplay instances

## Results for Task 1 pun detection EN



run ID	#	P	R	F <sub>1</sub>	A	P*	R*	F <sub>1</sub> *	A*
<i>Croland_EN_GPT3</i>	3,183	<b>100.00</b>	0.86	1.71	<b>74.80</b>	<b>100.00</b>	0.86	1.71	<b>74.80</b>
<i>LJGG_t5_large_easy_en</i>	3,183	42.73	71.94	<b>53.61</b>	68.36	42.73	71.94	<b>53.61</b>	68.36
<i>LJGG_t5_large_label_en</i>	3,183	25.41	<b>100.00</b>	40.53	25.41	25.41	<b>100.00</b>	40.53	25.41
<i>LJGG_t5_large_no_label_en</i>	3,183	25.41	<b>100.00</b>	40.53	25.41	25.41	<b>100.00</b>	40.53	25.41
<i>Les_miserables_fasttext</i>	3,183	25.78	80.96	39.11	35.94	25.78	80.96	39.11	35.94
<i>Les_miserables_random</i>	3,183	26.43	51.29	34.88	51.33	26.43	51.29	34.88	51.33
<i>Les_miserables_simplet5</i>	3,183	28.13	88.75	42.72	39.52	28.13	88.75	42.72	39.52
<i>MiCroGerk_EN_BLOOM</i>	13.00	8.33	0.12	0.24	74.26	8.33	<b>100.00</b>	15.38	15.38
<i>MiCroGerk_EN_FastText</i>	3,183	25.87	82.94	39.44	35.28	25.87	82.94	39.44	35.28
<i>MiCroGerk_EN_MLP</i>	3,183	29.04	72.92	41.54	47.84	29.04	72.92	41.54	47.84
<i>MiCroGerk_EN_NB</i>	3,183	25.98	95.42	40.84	29.75	25.98	95.42	40.84	29.75
<i>MiCroGerk_EN_Ridge</i>	3,183	26.74	85.16	40.70	36.94	26.74	85.16	40.70	36.94
<i>MiCroGerk_EN_SimpleT5</i>	3,183	30.75	83.06	44.88	48.16	30.75	83.06	44.88	48.16
<i>Smroltra_EN_FastText</i>	3,183	25.62	80.34	38.85	35.72	25.62	80.34	38.85	35.72
<i>Smroltra_Logistic-Regression</i>	3,183	26.14	86.15	40.11	34.62	26.14	86.15	40.11	34.62
<i>Smroltra_EN_MLP</i>	3,183	27.78	72.43	40.16	45.14	27.78	72.43	40.16	45.14
<i>Smroltra_EN_NBC</i>	3,183	26.12	95.55	41.02	30.19	26.12	95.55	41.02	30.19
<i>Smroltra_EN_SimpleT5</i>	3,183	31.97	83.68	46.27	50.61	31.97	83.68	46.27	50.61
<i>Smroltra_EN_TFIDF</i>	3,183	26.90	84.05	40.76	37.92	26.90	84.05	40.76	37.92
<i>TeamCAU_EN_A121</i>	40	27.58	0.98	1.90	74.17	27.58	80.00	41.02	0.43
<i>TeamCAU_EN_BLOOM</i>	40	30.00	0.37	0.73	74.45	30.00	30.00	30.00	65.00
<i>TeamCAU_EN_FastText</i>	3,183	25.71	80.84	39.02	35.78	25.71	80.84	39.02	35.78
<i>TeamCAU_EN_RandomForest</i>	3,183	25.69	83.43	39.28	34.46	25.69	83.43	39.28	34.46
<i>TeamCAU_EN_ST5</i>	3,183	26.99	93.32	41.87	34.15	26.99	93.32	41.87	34.15
<i>TeamCAU_EN_TFidfRidge</i>	3,183	26.74	85.16	40.70	36.94	26.74	85.16	40.70	36.94
<i>TheLangVerse_fasttext-MLP</i>	3,183	26.31	75.40	39.01	40.08	26.31	75.40	39.01	40.08
<i>ThePunDetectives_Fasttext</i>	3,183	26.07	80.22	39.35	37.16	26.07	80.22	39.35	37.16
<i>ThePunDetectives_NaiveBayes</i>	3,183	25.43	99.62	40.52	25.66	25.43	99.62	40.52	25.66
<i>ThePunDetectives_Ridge</i>	3,183	27.44	88.75	41.92	37.51	27.44	88.75	41.92	37.51
<i>ThePunDetectives_Roberta</i>	3,183	26.11	91.96	40.67	31.82	26.11	91.96	40.67	31.82
<i>ThePunDetectives_SimpleT5</i>	3,183	29.21	93.20	44.48	40.87	29.21	93.20	44.48	40.87
<i>UBO_SimpleT5</i>	3,183	36.51	85.53	51.18	58.52	36.51	85.53	51.18	58.52
<i>AKRaNLU_sentemb</i>	3,183	26.29	86.40	40.32	34.99	26.29	86.40	40.32	34.99
<i>AKRaNLU_seqclassification</i>	3,183	25.41	<b>100.00</b>	40.53	25.41	25.41	<b>100.00</b>	40.53	25.41
<i>Innsbruck_DS_backtranslation</i>	3,183	27.35	84.91	41.38	38.86	27.35	84.91	41.38	38.86
<i>Innsbruck_DS_r1</i>	3,183	27.32	86.89	41.57	37.92	27.32	86.89	41.57	37.92
<i>Innsbruck_DS_synonym</i>	3,183	27.15	86.89	41.37	37.41	27.15	86.89	41.37	37.41

## Results for wordplay detection FR



run ID	#	P	R	F <sub>1</sub>	A	P*	R*	F <sub>1</sub> *	A*
<i>Croland_FR_GPT3</i>	12,873	<b>100.00</b>	01.14	02.27	59.24	<b>100.00</b>	01.14	02.27	59.24
<i>LJGG_t5_large_easy_fr</i>	12,873	55.13	64.29	59.36	63.70	55.13	64.29	59.36	63.70
<i>LJGG_t5_large_label_fr</i>	12,873	41.23	<b>100.00</b>	58.39	41.23	41.23	<b>100.00</b>	58.39	41.23
<i>LJGG_t5_large_no_label_fr</i>	12,873	41.23	<b>100.00</b>	58.39	41.23	41.23	<b>100.00</b>	58.39	41.23
<i>Les_miserables_fasttext</i>	12,873	58.57	19.76	29.55	61.15	58.57	19.76	29.55	61.15
<i>Les_miserables_random</i>	12,873	41.14	49.81	45.06	49.92	41.14	49.81	45.06	49.92
<i>Les_miserables_simplet5</i>	12,873	59.72	74.88	<b>66.45</b>	68.82	59.72	74.88	<b>66.45</b>	68.82
<i>Smroltra_FR_FastText</i>	12,873	55.24	25.00	34.42	60.72	55.24	25.00	34.42	60.72
<i>Smroltra_FR_Logistic-Regression</i>	12,873	58.43	60.39	59.40	65.95	58.43	60.39	59.40	65.95
<i>Smroltra_FR_MLP</i>	12,873	56.49	62.88	59.52	64.73	56.49	62.88	59.52	64.73
<i>Smroltra_FR_NBC</i>	12,873	56.73	63.18	59.78	64.94	56.73	63.18	59.78	64.94
<i>Smroltra_FR_Random</i>	12,873	42.14	67.70	51.95	48.36	42.14	67.70	51.95	48.36
<i>Smroltra_FR_SimpleT5</i>	12,873	61.21	67.69	64.29	<b>68.99</b>	61.21	67.69	64.29	<b>68.99</b>
<i>Smroltra_FR_TFIDF</i>	12,873	58.77	62.09	60.38	66.41	58.77	62.09	60.38	66.41
<i>UBO_SimpleT5</i>	12,871	67.80	58.76	62.95	71.49	67.80	58.76	62.95	71.49
<i>AKRaNLU_sentemb</i>	12,873	41.18	73.88	52.88	45.71	41.18	73.88	52.88	45.71
<i>AKRaNLU_seqclassification</i>	12,873	41.23	<b>100.00</b>	58.39	41.23	41.23	<b>100.00</b>	58.39	41.23

## Results for Task 1 pun detection ES



run ID	#	P	R	F <sub>1</sub>	A	P*	R*	F <sub>1</sub> *	A*
<i>Croland_ES_GPT3</i>	2,241	<b>98.07</b>	05.35	10.15	<b>59.75</b>	<b>98.07</b>	05.35	10.15	59.75
<i>LJGG_t5_large_easy_es</i>	2,230	50.34	54.09	52.15	57.83	50.34	54.26	52.23	57.75
<i>LJGG_t5_large_label_es</i>	2,230	42.55	<b>99.68</b>	<b>59.64</b>	42.70	42.55	<b>100.00</b>	<b>59.70</b>	42.55
<i>LJGG_t5_large_no_label_es</i>	2,230	42.55	<b>99.68</b>	<b>59.64</b>	42.70	42.55	<b>100.00</b>	<b>59.70</b>	42.55
<i>Les_miserables_fasttext</i>	2,230	0.00	0.00	0.00	57.51	0.00	0.00	0.00	57.44
<i>Les_miserables_random</i>	2,230	43.43	51.78	47.24	50.87	43.43	51.94	47.31	50.76
<i>Les_miserables_simplet5</i>	2,230	51.10	17.01	25.53	57.83	51.10	17.07	25.59	57.75
<i>NLPalma_BERT</i>	2,230	55.94	40.54	47.01	61.17	55.94	40.67	47.10	61.12
<i>Smoltra_ES_FastText</i>	2,238	40.75	0.625	49.33	45.47	40.75	0.625	49.33	45.39
<i>Smoltra_ES_Logistic-Regression</i>	2,238	0.50	49.05	49.52	57.51	0.50	49.05	49.52	57.46
<i>Smoltra_ES_MLP</i>	2,238	55.45	44.32	49.27	<b>61.22</b>	55.45	44.32	49.27	61.17
<i>Smoltra_ES_NBC</i>	2,238	47.69	56.40	51.68	55.19	47.69	56.40	51.68	55.13
<i>Smoltra_ES_Random</i>	2,241	42.05	67.85	51.92	46.63	42.05	67.85	51.92	46.63
<i>Smoltra_ES_SimpleT5</i>	2,238	44.31	46.21	45.24	52.47	44.31	46.21	45.24	52.41
<i>Smoltra_ES_TFIDF</i>	2,238	53.34	46.11	49.46	59.97	53.34	46.11	49.46	59.91
<i>UBO_SimpleT5</i>	2,230	51.28	62.92	56.50	58.85	51.28	63.11	56.58	58.78
<i>AKRaNLU_sentemb</i>	2,230	41.39	72.26	52.63	44.75	41.39	72.49	52.70	44.61
<i>AKRaNLU_seqclassification</i>	2,230	42.55	<b>99.68</b>	<b>59.64</b>	42.70	42.55	<b>100.00</b>	<b>59.70</b>	42.55

# Conclusions on Task 1



- Homogeneous corpus in terms of vocabulary & length for pun detection in EN, FR, & ES
- 12 participants, 75 runs
  - Traditional classifiers & LLMs
- Pun detection is still a challenging task
  - maximum  $F_1 = 60\text{--}65\%$
- T5 trained on our corpus appeared to be most effective for all the languages
- Performance of methods varied depending on implementation, fine-tuning, the specific prompts used, etc.
- Overfitting issues
- Incomplete runs (LLMs)
  - time → importance of efficiency
  - token limits

## Task 2: Pun Location & Interpretation



- *Pun location* is a finer-grained task, where the goal is to identify which words carry the double meaning in a text known *a priori* to contain a pun.
  - In *pun interpretation*, systems have to indicate the two meanings of the pun.
    - lemmatised word sets: synonyms or hypernyms of the two words involved in the pun
  - **Data**: manually annotated corpus from Task 1 in EN and FR
  - **Evaluation**: average scores for each of senses in terms of P, R, F1, Acc
- (3) When the church bought gas for their annual barbecue, proceeds went from the sacred to the **propane**. - {*gas, fuel*} & {*profane*}
- (4) I used to be a banker but I lost **interest**. - {*involvement*} & {*fixed charge, fixed cost, fixed costs*}

# Dataset statistics for Task 2



Language	Train	Test
<i>English</i>	2,315	1,205
<i>French</i>	2,000	4,655
<i>Spanish</i>	876	960

# Results for Task 2.1 (pun location)



run ID	EN			FR			ES		
	#	A	A*	#	A	A*	#	A	A*
<i>Croland_GPT3</i>	19	0.41	26.31	61	0.20	18.03	51	1.77	33.33
<i>Les_miserables_random</i>	1205	8.87	8.87	4655	4.37	4.98	960	6.14	6.14
<i>Les_miserables_simpleT5</i>	1205	76.18	76.18	4655	39.92	45.49	960	55.41	55.41
<i>Les_miserables_word</i>	1205	49.54	49.54	4655	28.67	32.67	960	51.56	51.56
<i>Smroltra_BLOOM</i>	32	1.74	65.62	65	0.41	33.84	57	2.60	43.85
<i>Smroltra_GPT3</i>	32	2.15	81.25	65	0.56	46.15	57	5.20	87.71
<i>Smroltra_SimpleT5</i>	1205	79.50	79.50	4655	39.86	45.43	960	<b>82.81</b>	82.81
<i>Smroltra_SpaCy</i>	1205	44.48	44.48	4655	0.00	0.00	960	24.16	24.16
<i>UBO_SimpleT5</i>	1205	77.67	77.67	4655	40.39	46.03	960	57.70	57.70
<i>AKRaNLU_x</i>	1205	77.51	77.51	4655	40.56	46.22	960	54.27	54.27
<i>AKRaNLU_y</i>	1205	79.17	79.17	4655	<b>41.35</b>	47.13	960	56.14	56.14
<i>TeamCAU_AI21</i>	32	1.16	43.75	—	—	—	—	—	—
<i>TeamCAU_BLOOM</i>	32	1.24	46.87	—	—	—	—	—	—
<i>TeamCAU_ST5</i>	1205	80.66	80.66	—	—	—	—	—	—
<i>ThePunDetectives_Fasttext</i>	1205	5.06	5.06	—	—	—	—	—	—
<i>ThePunDetectives_NaiveBayes</i>	1205	2.07	2.07	—	—	—	—	—	—
<i>ThePunDetectives_Ridge</i>	1205	50.20	50.20	—	—	—	—	—	—
<i>ThePunDetectives_SimpleT5</i>	1205	80.41	80.41	—	—	—	—	—	—
<i>ThePunDetectives_SimpleTransformersT5</i>	1205	<b>83.15</b>	83.15	—	—	—	—	—	—
<i>MiCroGerk_AI21</i>	17	1.32	94.11	—	—	—	—	—	—
<i>MiCroGerk_BLOOM</i>	17	0.99	70.58	—	—	—	—	—	—
<i>MiCroGerk_OpenAI</i>	17	1.24	88.23	—	—	—	—	—	—
<i>MiCroGerk_SimpleT5</i>	1205	79.91	79.91	—	—	—	—	—	—
<i>MiCroGerk_lastWord</i>	1205	54.43	54.43	—	—	—	—	—	—
<i>MiCroGerk_random</i>	1205	13.94	13.94	—	—	—	—	—	—



## Results for Task 2.2 EN (interpretation)



run	count	score	part_score
<i>C&amp;O_task_2.2_Chat GPT</i>	92	5.45%	<b>70.65%</b>
<i>Croland_task_2_EN_GPT3</i>	29	0.08%	3.45%
<i>Les_miserables_simplet5</i>	1,192	<b>47.40%</b>	47.40%
<i>MiCroGerk_task_2_EN_AI21</i>	11	0.46%	50.00%
<i>MiCroGerk_task_2_EN_BLOOM</i>	2	0.04%	25.00%
<i>MiCroGerk_task_2_EN_OpenAI</i>	11	0.34%	36.36%
<i>MiCroGerk_task_2_EN_SimpleT5</i>	39	1.59%	48.72%
<i>Smroltra_task_2_Bloom</i>	32	0.59%	21.88%
<i>Smroltra_task_2_GPT3</i>	32	0.59%	21.88%
<i>Smroltra_task_2_GPT3_WN</i>	32	1.09%	40.63%
<i>Smroltra_task_2_SimpleT5_WN</i>	1192	41.44%	41.44%
<i>Smroltra_task_2_bloom_WN</i>	32	0.80%	29.69%
<i>Smroltra_task_2_spaCy_WN</i>	1,192	19.76%	19.76%
<i>UBO_task_2.2_SimpleT5</i>	1,192	46.85%	46.85%
<i>akranlu_task_2.2_sentembwordnet</i>	1,192	39.77%	39.77%

# Conclusions on Task 2



- Semantic annotation for wordplay in English and French
  - a new corpus for pun location in Spanish
  - 9 teams, 64 runs
    - similar methods, but different results
- Wordplay location is still a challenge for LLMs
  - results for pun location in FR are quite low (different data creation procedures?)
    - Some of the puns in EN & ES might have been “known” by LLMs
    - French data was novel and largely constructed by us
  - Overfitting
- Pun interpretation is quite a challenging task due to the inherent ambiguity and creativity of puns
  - We evaluated pun interpretation only for EN
  - Many partial runs

# Task 3: Pun Translation



- The goal of this task is to translate English punning jokes into French and Spanish:
  - contain an instance of wordplay
  - preserve wordplay meaning
- **Data:** English puns from Task 1 manually translated into French and Spanish
- **Evaluation:** Experts

# Task 3 dataset statistics



Language	Train		Test	
	target	source	target	source
<i>French</i>	5,838	1,405	6,590	1,197
<i>Spanish</i>	644	217	5,727	544

## Results for pun translation EN→FR



run ID	#E	#T	#M	%M	#W	%W	#S	%S	%R
<i>Croland_GPT3</i>	16	28	4	25	0	0	0	0	0
<i>LJGG_Google_Translator_auto</i>	1,076	1,197	580	53	67	6	63	5	5
<i>LJGG_fr_mt5_base_auto</i>	2	1,197	2	100	1	50	1	50	0
<i>LJGG_fr_mt5_base_no_label_auto</i>	1	1,197	1	100	0	0	0	0	0
<i>LJGG_fr_t5_large_auto</i>	90	1,197	24	26	2	2	2	2	0
<i>LJGG_fr_t5_large_no_label_auto</i>	140	1,197	80	57	15	10	15	10	1
<i>Smroltra_BLOOM</i>	31	32	8	25	0	0	0	0	0
<i>Smroltra_EasyNMT-Opus</i>	786	1,197	427	54	58	7	56	7	4
<i>Smroltra_EasyNMT-mbart</i>	1,139	1,197	613	53	68	5	64	5	5
<i>Smroltra_GPT3</i>	30	32	8	26	0	0	0	0	0
<i>Smroltra_GoogleTranslation</i>	1,109	1,197	602	54	71	6	67	6	5
<i>Smroltra_SimpleT5</i>	1,043	1,197	562	53	66	6	65	6	5
<i>TeamCAU_AI21</i>	30	32	8	26	0	0	0	0	0
<i>TeamCAU_BLOOM</i>	32	32	8	25	0	0	0	0	0
<i>TeamCAU_ST5</i>	1,090	1,197	577	52	71	6	69	6	5
<i>TheLangVerse_j2-grande-finetuned</i>	1,176	1,197	636	54	76	6	72	6	6
<i>ThePunDetectives_M2M100</i>	13	340	9	69	2	15	2	15	0
<i>ThePunDetectives_OpusMT</i>	183	340	92	50	19	10	17	9	1
<i>UBO_SimpleT5</i>	73	1,195	47	64	5	6	5	6	0
<i>UBO_SimpleT5_x</i>	1,148	1,195	616	53	71	6	67	5	5
<i>UBO_SimpleT5_y</i>	791	1,194	429	54	61	7	59	7	5

## Results for pun translation EN→ES



run ID	#E	#T	#M	%M	#W	%W	#S	%S	%R
<i>Croland_ENESGPT3</i>	45	47	9	20.00	3	6.66	3	6.66	0
<i>LJGG_es_mt5_base_auto</i>	34	544	16	47.05	5	14.70	5	14.70	0
<i>LJGG_es_mt5_base_no_label_auto</i>	34	544	16	47.05	5	14.70	5	14.70	0
<i>LJGG_es_t5_large_auto</i>	34	544	16	47.05	5	14.70	5	14.70	0
<i>LJGG_es_t5_large_no_label_auto</i>	34	544	16	47.05	5	14.70	5	14.70	0
<i>LJGG_GoogleTranslatorENESauto</i>	544	544	274	50.36	106	19.48	99	18.19	18
<i>NLPalma_BLOOMZ_x</i>	359	359	215	59.88	85	23.67	80	22.28	14
<i>NLPalma_BLOOMZ_y</i>	359	359	215	59.88	85	23.67	80	22.28	14
<i>Smroltra_EasyNMT-Opus</i>	529	544	263	49.71	100	18.90	93	17.58	17
<i>Smroltra_GoogleTranslation</i>	532	544	267	50.18	103	19.36	96	18.04	17
<i>Smroltra_SimpleT5</i>	531	544	265	49.90	101	19.02	94	17.70	17
<i>Smroltra_ENESBLOOM</i>	45	47	8	17.77	2	4.44	2	4.44	0
<i>TheLangVerse_j2-grande-finetuned</i>	415	544	200	48.19	70	16.86	65	15.66	11
<i>ThePunDetectives_M2M100</i>	33	430	16	48.48	7	21.21	7	21.21	1
<i>ThePunDetectives_ENESOpusMT</i>	428	430	208	48.59	71	16.58	66	15.42	12
<i>MiCroGerk_OpenAI</i>	6	17	3	0.5	1	16.66	1	16.66	0
<i>MiCroGerk_mbart50_m2m_x</i>	543	544	274	50.46	106	19.52	99	18.23	18
<i>MiCroGerk_AI21_x</i>	1	17	1	1	0	0	0	0	0
<i>MiCroGerk_m2m_100_418M</i>	43	544	23	53.48	11	25.58	11	25.58	2
<i>MiCroGerk_SimpleT5</i>	5	544	4	0.8	3	0.6	3	0.6	0

# Conclusions on Task 3



- New parallel corpora of wordplay EN-FR & EN-ES
- Manual evaluation → augmented data for wordplay detection
- Success rate of wordplay translation is extremely low for both language pairs
  - Max of 6% over the total evaluated *test* set for FR
  - Max of 17% over the total evaluated *train* set for FR
  - Max of 18% over the total evaluated *test* set for ES

# JOKER program (ROOM 3)



- Wed 20th Sept (TODAY) - CONFERENCE Session 5 NLP (Room 1)
  - *Humour Translation with Transformers*, Farhan Dhanani, Muhammad Rafi and Muhammad Atif Tahir (Best of the Labs 2022)
- Thu 21th Sept (TOMORROW)
  - JOKER Session 1 at 09:30 - 11:00
    - Invited talk by Jussi Karlgren (Silo AI) on *How to evaluate delight and diversion?*
    - JOKER Track overview papers
  - JOKER Session 2 at 11:30 - 13:00
    - Participant presentations
  - JOKER Session 3 at 14:00 - 15:30
    - Invited talk by Alexander Libov (Amazon) on *Characterizing Playful Requests to Virtual Assistants and Evaluating Playful Response Generation*
    - Round table and JOKER 2024 discussion





*Thank you !  
Participate in our track !*

Website : <https://joker-project.com><sup>1</sup>

E-mail : [contact@joker-project.com](mailto:contact@joker-project.com)

Twitter : [https://twitter.com/joker\\_research](https://twitter.com/joker_research)

Google group : <https://groups.google.com/g/joker-project>

---

<sup>1</sup>This project has received a government grant managed by the National Research Agency under the program "Investissements d'avenir" integrated into France 2030, with the Reference ANR-19-GURE-0001. JOKER is supported by *La Maison des sciences de l'homme en Bretagne*.