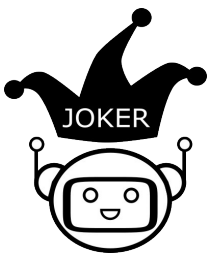
CLEF 2025 JOKER Track

Humor Detection, Search, and Translation

Liana Ermakova **Anne-Gwenn Bosser**
Tristan Miller **Ricardo Campos**



CLEF, September 9, 2025, Madrid, Spain

JOKER Track Motivation

- Official Motivation:
  - State-of-the-art AI, NLP, and IR models are not able to deal with humour or other non-literal meaning aspects of texts
  - Wordplay, which can refer to the surface structure or orthography of a word or its pronunciation, is a challenge for AI
  - Word surface aspects are not captured in the deep semantic embedding in AI models
  - These also cannot be captured by current pre-training models based on next-word prediction objectives
- Motivation of students in the track:
  - *Serious research* on *funny topics*?
  - I can work on humor and jokes AND my thesis at the same time?

JOKER 2025 Tasks

1. **Humor-aware Information Retrieval**: *retrieve short humorous texts for a query.*
2. ***Wordplay Translation***: *translate puns from English to French.*
3. ***Onomastic Wordplay Translation***: *translate humourous namesfrom English to French.*

Statistics

Team	Task 1		Task 2	Task 3	Total
	EN	PT			
alecs			8		8
arampageos	3	9	12	15	39
cryptix	3		1		4
fhelms	4				4
igoranchik	3	2	13	2	20
kamps	4	4	2		10
mariapazr20				1	1
pjmathematician	7	4	4	3	18
rasion	2	2			4
rdtaylorjr			4		4
sarath_kumar	1		5	1	7
tanishc228	14				14
verbanex			3		3
baselines				1	1
Total	41	21	52	22	136

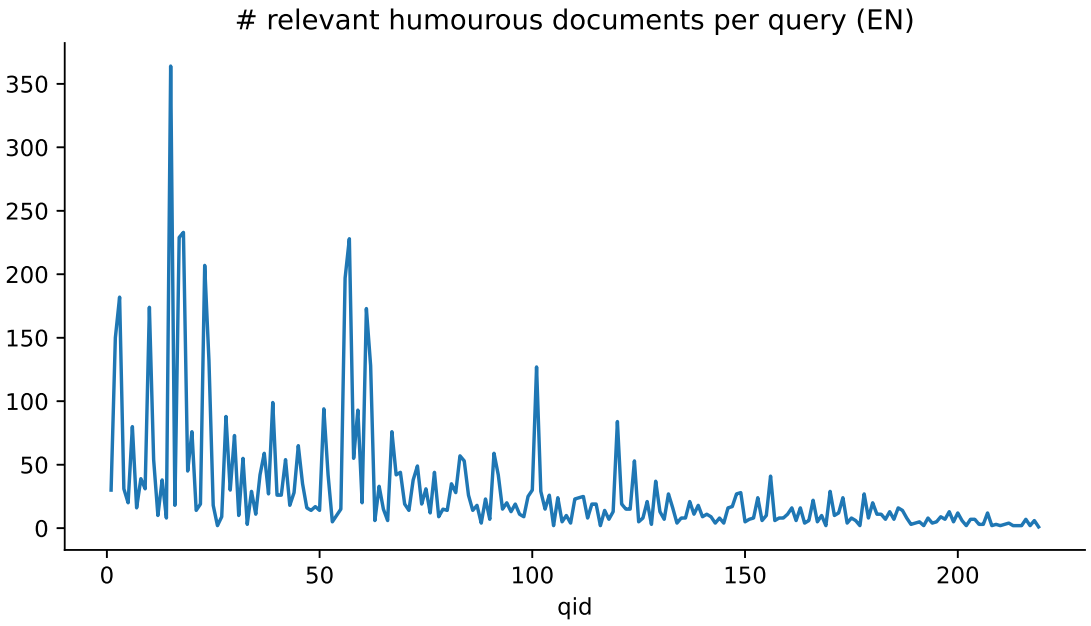
Task 1: Humor-aware IR

- *Task 1: Retrieve short humorous texts for a query*
  - Retrieving short humorous texts from a document collection
- Use case: to search for a joke on a specific topic
- *Main JOKER IR Task since 2024*
- *219 queries with relevance judgments (11 queries for Train)*
- *77,658 English documents*
  - *5,198 humorous texts (JOKER 2024 + new jokes)*
- *New data*
  - *New manually created jokes*
  - *New positive generated examples*
  - *New negative examples*
- *Evaluation: traditional IR metrics (MAP, NDCG, ...)*
- *+ Data in Portuguese*

EN Data Source Statistics

source	non-humorous	humorous	total
Bard	36	4	40
Claude	0	74	74
ChatGPT	149	381	530
JOKER	4,954	3,507	8,461
Llama-2	12,523	0	12,523
Phi-3 Mini	8,204	0	8,204
manual	2	247	249
translations	985	0	985
Wikipedia	46,592	0	46,592
total	72,460	5,198	77,658

relevant humourous documents per query in EN



count	219
mean	30
std	48
min	1
25%	7
50%	15
75%	29
max	364

Document Examples


```
[
  {
    "docid": "1",
    "text":  "Good laws have sprung from bad
       customs."
  },
  {
    "docid": "2",
    "text":  "The musical score to Topsyturveydom
        does not survive, but amateur productions
        in recent decades have used newly
       composed scores or performed the work as a
        non-musical play."
  },
  {
    "docid": "3",
    "text": "The organic compound primarily
       responsible for the characteristic odor of
        musk is muscone."
  },
  {
    "docid": "51135",
    "text":  "I've inherited a fortune, said Tom,
        willfully."
  },
  {
    "docid": "591",
    "text":  "My name is Will, I'm a lawyer."
  }
]
```

Results Task 1 EN (Test)

Run ID	#ret	#rel	map	gmap	p@r	mrr	p@5	p@10	ndcg@5
pjmathematician_Q14-Q8-Q32	207K	3007	35.01	24.65	38.68	79.04	54.88	40.63	60.80
pjmathematician_Q14-Q8-Q14	207K	2954	34.86	24.31	39.07	78.74	54.49	40.29	60.59
pjmathematician_Q32-Q8-Q14	207K	2932	34.38	23.98	38.98	80.78	54.49	41.01	60.94
pjmathematician_Q32-Q8-Q32	207K	3011	32.91	23.03	36.99	76.20	50.34	39.95	55.98
pjmathematician_Q14-Q8-R	207K	2835	23.88	16.03	27.10	64.27	36.62	28.70	41.23
UAms_RM3RoBERTa_drop60	82K	1448	16.72	7.04	23.05	54.46	30.82	23.09	1.52
Rasion_SenTransF+Roberta	4K	811	16.21	NaN	20.59	64.93	35.92	24.47	41.34
Rasion_SenTransF+Roberta	53K	1475	15.79	5.53	20.21	55.94	30.82	22.75	34.02
Cryptix_SBERT	207K	1914	15.07	5.52	19.44	56.94	28.70	20.97	33.46
UAms_RM3	207K	1864	15.02	7.22	19.53	40.87	24.35	20.00	25.66
UAms_RM3RoBERTa	186K	1798	14.94	7.16	19.56	42.47	25.51	19.61	26.76
CCC_Ensemble_ColBERT_RM3	103K	1967	14.15	7.44	16.29	33.69	16.71	17.73	18.17
CCC_Ensemble	206K	2050	14.03	7.12	17.01	40.33	20.48	18.65	22.55
UAms_RM3	207K	1872	12.16	5.77	15.36	33.30	18.55	18.07	19.57
UAms_en_bm25	41K	1884	11.91	5.64	12.23	26.28	12.95	12.71	14.00
CCC_TFIDF_Rerank	100K	1764	11.26	NaN	15.25	40.59	20.59	17.07	22.04
UAms_Anserini	207K	2134	10.76	5.35	10.56	25.03	11.88	12.22	12.37
UAms_en_rm3	207K	2134	10.76	5.35	10.56	25.03	11.88	12.22	12.37
CCC_ColBERT_Enhanced	207K	1879	9.93	5.17	12.47	33.35	15.75	14.15	16.53
CCC_XLM_R_Rerank	207K	2227	9.66	5.03	9.83	36.41	13.82	11.98	16.19
CCC_ColBERT_Enhanced	207K	1418	6.69	2.06	9.34	31.24	13.43	11.40	14.55
CCC_XLM_R_Rerank	10K	918	6.30	2.56	9.65	23.88	9.18	9.08	19.10
CCC_ColBERT_Enhanced	207K	1367	6.21	1.82	8.67	28.52	11.30	10.34	12.42
CCC_Advanced_Ensemble_LTR	165K	2122	6.20	3.65	6.38	11.54	2.90	5.89	2.67
CCC_TFIDF	207K	1321	5.79	1.56	8.41	25.29	9.47	9.03	10.50
CCC_Ensemble_ColBERT_RM3	103K	1904	5.44	2.24	6.52	21.03	8.12	6.96	8.77
CCC_TF-IDF_Ens_ColBERT_RM3	103K	1922	5.31	2.22	5.98	20.16	7.15	6.47	7.90
Skommarkhos_BM25_E5_MiniLM	207K	2182	5.02	2.98	3.03	6.47	0.87	3.24	0.65
UAms_en_bm25_CE1K	41K	1884	4.88	2.60	2.47	5.68	0.48	2.75	0.38
UAms_en_rm3_CE1K	207K	2134	4.78	2.67	2.32	5.48	0.39	2.51	0.27
UAms_Anserini	10K	849	4.76	1.41	3.92	7.67	0.87	3.67	0.75
cryptix_crossencoder	20K	999	3.78	1.55	2.43	5.64	0.48	2.51	0.34
CCC_Ensemble_RoBERTa_RM3	41K	1718	3.33	1.69	4.02	11.01	3.77	3.82	23.65
Skommarkhos_BM25_E5_MiniLM	20K	517	2.49	0.49	2.62	6.15	0.77	3.19	0.57
CCC_pipeline	5K	211	2.43	0.05	3.93	11.81	4.54	4.15	0.80
team_reranker_EN	20K	824	1.38	0.23	2.40	6.36	2.13	2.03	2.05
yourteam_xlm_roberta_large	20K	271	0.42	0.02	1.29	4.71	1.35	1.50	1.30
duth_xanthi_en	20K	62	0.04	0.00	0.21	0.75	0.10	0.10	0.08
cryptix_crossencoder	414K	336	0.02	0.00	0.01	0.08	0.00	0.00	0.00

Results Task 1 EN (Train)

Run ID	#ret	#rel	map	gmap	p@r	mrr	p@5	p@10	ndcg@5
Rasion_SenTransF+Roberta	1K	642	59.28	56.10	61.41	69.72	60.00	52.50	60.87
pjmathematician_Q32-Q8-Q14	12K	395	48.10	35.02	47.15	78.47	66.67	53.33	67.55
pjmathematician_Q32-Q8-Q32	12K	439	45.91	32.72	45.17	73.06	60.00	49.17	59.80
pjmathematician_Q14-Q8-Q14	12K	411	44.90	30.77	47.98	70.74	56.67	50.83	56.99
pjmathematician_Q14-Q8-Q32	12K	440	44.28	30.38	45.55	69.63	51.67	50.83	53.27
Rasion_SenTransF+Roberta	1K	596	42.91	39.31	45.86	72.50	45.00	46.67	48.37
Cryptix_SBERT	12K	571	39.21	27.13	40.63	79.32	56.67	42.50	59.59
pjmathematician_Q14-Q8-R	12K	532	32.35	20.79	36.25	53.19	40.00	36.67	38.45
CCC_Ensemble	11K	477	18.54	4.46	18.89	47.78	21.67	25.83	24.70
CCC_XLM_R_Rerank	12K	538	18.52	3.67	18.45	36.79	18.33	15.00	20.91
CCC_Ensemble_ColBERT_RM3	6K	366	17.57	4.77	19.48	36.52	23.33	20.00	22.86
CCC_TFIDF_Rerank	5K	359	17.14	4.10	25.53	52.51	25.00	24.17	27.53
UAms_RM3	12K	465	17.09	2.63	16.62	33.53	20.00	16.67	21.79
UAms_RM3RoBERTa_drop60	4K	252	16.33	1.48	19.89	49.65	31.67	27.50	6.75
CCC_ColBERT_Enhanced	12K	473	15.99	3.17	17.19	31.34	21.67	16.67	20.88
UAms_RM3RoBERTa	10K	459	15.69	2.17	19.62	39.93	23.33	22.50	21.20
UAms_RM3	12K	459	15.53	2.08	19.44	39.07	21.67	21.67	20.11
UAms_Anserini	12K	476	15.51	3.39	13.61	29.91	16.67	17.50	16.86
UAms_en_rm3	12K	476	15.51	3.39	13.61	29.91	16.67	17.50	16.86
UAms_en_bm25	4K	471	14.22	3.31	13.71	21.40	6.67	13.33	7.24
CCC_ColBERT_Enhanced	12K	391	10.97	0.53	14.48	32.95	15.00	16.67	16.62
CCC_TFIDF	12K	383	10.16	0.49	12.31	34.58	15.00	14.17	17.11
CCC_ColBERT_Enhanced	12K	382	9.97	0.47	11.84	37.55	16.67	14.17	18.93
CCC_Ensemble_ColBERT_RM3	6K	371	9.79	1.94	10.41	21.90	13.33	12.50	12.62
CCC_TF-IDF_Ens_ColBERT_RM3	6K	370	9.59	1.84	10.77	19.15	13.33	12.50	12.06
CCC_Advanced_Ensemble_LTR	9K	487	9.46	2.17	7.65	8.18	3.33	6.67	2.31
Skommarkhos_BM25_E5_MiniLM	12K	442	9.29	5.13	6.65	9.03	6.67	6.67	5.35
UAms_en_rm3_CE1K	12K	476	8.63	1.76	7.17	8.24	5.00	6.67	4.04
UAms_en_bm25_CE1K	4K	471	8.59	1.77	7.63	8.24	5.00	6.67	4.04
CCC_XLM_R_Rerank	600	88	6.04	1.36	6.55	18.71	10.00	9.17	17.14
UAms_Anserini	600	74	5.31	0.60	4.83	12.67	8.33	8.33	6.97
CCC_Ensemble_RoBERTa_RM3	2K	207	4.94	1.46	5.19	20.98	15.00	12.50	29.57
cryptix_crossencoder	1K	122	4.45	1.08	3.38	7.79	5.00	4.17	4.09
Skommarkhos_BM25_E5_MiniLM	1K	62	2.59	0.71	2.68	8.64	5.00	4.17	4.29
team_reranker_EN	1K	46	1.37	0.28	2.42	7.90	5.00	3.33	4.09
yourteam_xlm_roberta_large	1K	16	0.95	0.04	2.66	10.07	5.00	3.33	4.04
CCC_pipeline	300	17	0.31	0.01	2.05	8.20	1.67	4.17	2.31
duth_xanthi_en	1K	9	0.04	0.01	0.54	4.78	1.67	1.67	1.78
cryptix_crossencoder	24K	18	0.01	0.00	0.02	0.09	0.00	0.00	0.00

Task 1: Humor-aware IR in PT

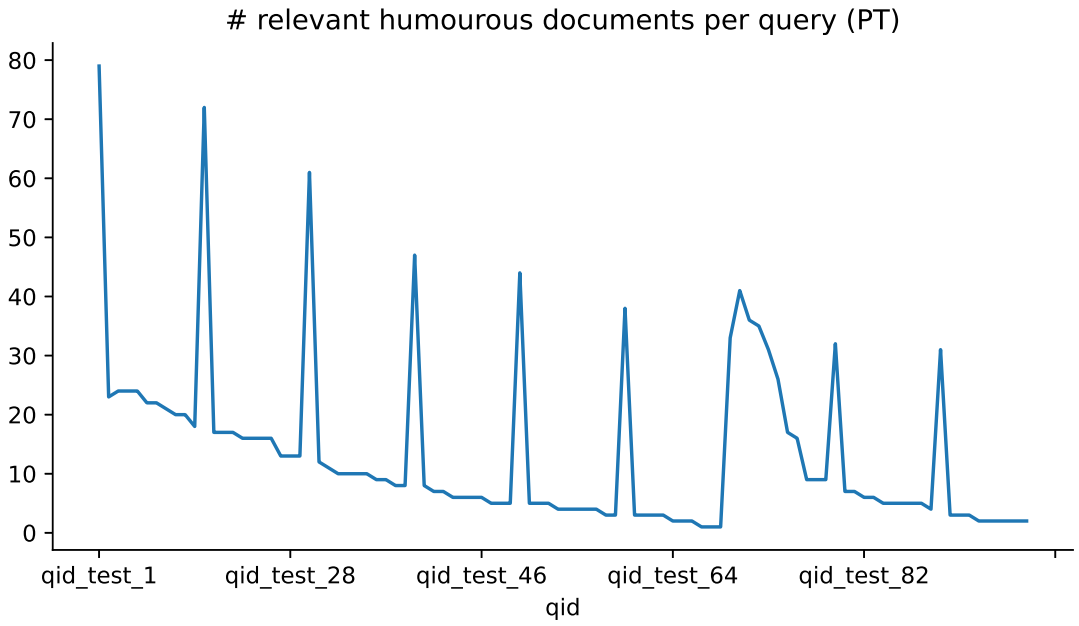
- New Portuguese Collection: 45,126 European Portuguese documents
  - 1199 humorous texts (660 texts translated from last year's EN collection + 539 new wordplay)
  - 2899 texts generated using ChatGPT3.5 Turbo
  - 41,028 sentences from Wikipedia extracts
  - 98 queries
- Curation and Translation Process:
  - To ensure consistency with European Portuguese (Pt-PT), PtVId[1] model was used to classify each text as Pt-PT or Brazilian Portuguese (Pt-BR).
  - Texts identified as Pt-BR were translated into Pt-PT using ChatGPT-4o-mini.
  - A final manual curation step was performed to verify translation quality and accuracy.

[1] Sousa, H. et al., Enhancing Portuguese Variety Identification with Cross-Domain Approaches, AAAI'25, 2025.

PT Data Source Statistics

source	#
chatgpt-3.5turbo	1615
joker	972
joker-2	227
wikipedia	15720
wikipedia-non-relevant	26592

relevant humourous documents per query in PT



count	98
mean	14
std	15
min	1
25%	4
50%	8
75%	17
max	79

Results Task 1 PT (Test)

Run ID	#ret	#rel	map	gmap	p@r	mrr	p@5	p@10	ndcg@5
pjmathematician_Q32-Q4-R	69000	932	42.21	30.78	42.01	69.07	43.77	34.35	42.14
pjmathematician_Q14-Q4-R	69000	938	42.17	30.81	41.65	68.98	43.77	34.49	51.69
Rasion_SenTransF+Roberta	69000	905	40.51	28.90	40.17	66.57	44.93	38.41	50.15
Rasion_SenTransF+Roberta	62576	904	40.51	28.90	40.17	66.57	44.93	38.41	50.12
UAms_pt_bm25	12856	229	7.89	0.19	5.96	9.83	5.22	6.09	5.13
Skommarkhos_BM25_E5_MiniLM	69000	503	7.42	1.65	5.74	11.91	6.38	6.23	6.44
pjmathematician_Q06-gist	69000	562	6.95	1.75	4.99	11.20	5.51	6.38	5.46
Skommarkhos_BM25_E5_MiniLM	6900	228	6.90	0.28	5.58	12.65	5.22	5.94	5.35
results_pt_pt_finetuned	6900	199	6.71	0.41	6.74	20.21	7.54	7.10	9.29
UAms_pt_rm3	67994	262	6.54	0.25	5.91	9.51	4.64	5.65	4.47
myteam_BERT	69000	496	6.13	1.26	6.38	19.54	8.12	6.38	8.78
duth_xanthi_pt	6900	225	5.95	0.37	6.76	15.65	7.54	8.41	7.03
pjmathematician_Q06-gist-exp32	69000	512	4.91	1.35	2.92	7.15	2.61	3.48	2.73
UAms_pt_rm3_CE1K	67994	262	4.16	0.19	2.47	5.20	1.45	3.19	1.34
UAms_pt_bm25_CE1K	12856	229	3.84	0.12	1.99	4.47	1.16	3.04	0.91
team_xlmr_PT	6900	133	2.96	0.11	5.33	12.03	4.64	5.94	4.57
results_pt_large_pt_finetuned	6900	65	0.31	0.01	0.02	0.73	0.00	0.00	1.72
yourteam_pt_zeroshot	6900	46	0.27	0.01	0.28	1.13	0.00	0.29	0.00
xlm-roberta-triplet-pt	6900	28	0.22	0.00	0.33	2.48	0.58	0.43	0.70

Results Task 1 PT (Train)

Run ID	#ret	#rel	map	gmap	p@r	mrr	p@5	p@10	ndcg@5
Skommarkhos_BM25_E5_MiniLM	29000	219	8.26	2.91	4.70	8.38	4.14	5.52	4.04
UAms_pt_bm25_CE1K	4474	115	4.69	0.24	3.00	4.11	2.07	3.10	1.54
results_pt_pt_finetuned	2900	100	9.84	0.55	10.56	19.19	8.28	6.90	10.33
duth_xanthi_pt	2900	107	5.86	0.46	7.07	15.95	5.52	5.52	7.83
yourteam_pt_zeroshot	2900	25	0.39	0.02	0.27	1.52	0.00	0.00	0.00
pjmathematician_Q32-Q4-R	29000	286	39.20	24.77	35.86	60.49	36.55	26.90	38.88
Rasion_SenTransF+Roberta	26477	296	49.00	29.84	49.37	65.47	42.07	32.07	52.24
yourteam_xlm-roberta-en	2900	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
pjmathematician_Q06-gist-exp32	29000	268	4.75	2.56	2.38	6.32	2.07	2.76	1.59
myteam_BERT	29000	191	8.13	2.11	7.21	18.23	7.59	4.83	9.63
xlm-roberta-triplet-pt	2900	6	0.08	0.00	0.11	0.68	0.00	0.00	0.00
results_pt_large_pt_finetuned	2900	47	0.61	0.01	0.27	4.32	0.69	0.34	5.28
pjmathematician_Q14-Q4-R	29000	278	39.46	24.51	36.17	60.41	36.55	26.55	45.40
Skommarkhos_BM25_E5_MiniLM	2900	116	7.62	0.51	3.92	9.81	3.45	6.21	3.47
pjmathematician_Q06-gist	29000	263	8.41	4.33	7.53	16.05	6.90	5.86	8.02
UAms_pt_bm25	4474	115	16.24	0.61	14.65	15.79	9.66	9.66	12.19
UAms_pt_rm3_CE1K	28905	124	4.68	0.39	3.00	4.12	2.07	3.10	1.54
team_xlmr_PT	2900	71	5.08	0.13	4.67	10.48	4.83	6.21	4.68
Rasion_SenTransF+Roberta	29000	297	49.00	29.84	49.37	65.47	42.07	32.07	52.24
yourteam_xlm-roberta-pt	2900	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UAms_pt_rm3	28905	124	8.86	0.64	7.00	9.25	4.83	6.55	4.85

Conclusions on Task 1

- EN: 77,658 documents (5,198 humorous) + 219 queries with relevance judgments
- PT: 45,126 documents (1,199 humorous) + 98 queries
- 9 teams submitted 41 runs for EN
- 5 teams submitted 21 runs for PT
- Diverse methods (TF–IDF, BM25, RM3, cross-encoders with and without filtering, LLMs etc.)
- Best results: Qwen model for retrieval and filtering and dense retrieval and transformer-based detection of humorous texts
- Nearly triple last year's top score
- Dataset's core properties have remained stable and RM3 and BM25 baselines showed stable performance

Task 2: Wordplay Translation

- *Task 2: Translate Puns from English to French and preserve*:
  - wordplay form
  - wordplay meaning
- **Train data:** 5,838 manual FR translations of 1,405 EN puns
- **Test data:** 2,615 new manual FR translations of 1,682 EN puns
- *Main JOKER NLP Task, continues from 2022-2024*

```
{
    "id_en":"en_1",
    "en":"I used to be a banker but I lost
       interest",
    "fr":"J'ai été banquier mais j'en ai
       perdu tout l'intérêt."
}
```

Evaluation

- **BLEU**
- **BERTScore** precision, recall, & F_1 over all references
- **Pun location–based evaluation**: # & % of words or phrases with multiple meanings (pun locations) in runs matching those in references
- **Manual evaluation**: human assessments of 1,297 French translations of 50 distinct source English puns in terms of meaning preservation and the presence of wordplay

Histogram of # of references & locations per EN pun

Train			Test		
#	references	locations	#	references	locations
1	396	578	1	1,252	1,382
2	237	254	2	172	220
3	234	192	3	133	68
4	169	110	4	53	10
5	88	70	5	39	1
6	55	58	6	22	1
7	42	31	7	8	—
8	27	15	8	2	—
9	25	15	9	1	—
10	12	11			
11	13	20			
12	20	12			
13	7	7			
14	8	6			
15	12	8			
16	10	7			
17	6	2			
18	8	3			
19	7	4			
20	6	2			
>20	11	—			

BLEU Scores on Test Data

run ID	Score	$n = 1$	$n = 2$	$n = 3$	$n = 4$
Skommarkhos_Lucie_SFT	43.33	65.05	46.98	37.59	30.67
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v4	43.20	64.73	46.74	37.50	30.69
UvA_finetunedNLLB-1.3B	42.55	64.74	46.26	36.70	29.83
Skommarkhos_Lucie_SFT_ARPO	42.48	63.76	45.92	36.86	30.17
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v8	42.26	64.44	46.13	36.47	29.42
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a5	42.15	63.33	45.50	36.56	29.95
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a1	42.14	63.38	45.53	36.54	29.90
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a7	42.14	63.37	45.55	36.56	29.87
Skommarkhos_Lucie-7B-Instruct-v1.1	42.14	63.43	45.54	36.51	29.88
Skommarkhos_Lucie-7B-Instruct-v1.1	42.12	63.41	45.54	36.50	29.86
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a11	42.11	63.33	45.46	36.51	29.91
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_arpo_a19	42.00	63.29	45.41	36.40	29.74
UvA_finetunedNLLB-1.3B&finetunedroBERTa	41.80	63.86	45.49	36.01	29.17
UvA_finetunedMarianMT	41.19	63.37	44.74	35.31	28.76
duth_hybrid_fusion	41.11	63.45	44.62	35.17	28.70
yourteamid_marianmt_pun_postedit	41.01	63.40	44.52	35.07	28.58
duth_xanthi_helsinki	41.01	63.40	44.52	35.07	28.58
Cryptix	41.01	63.40	44.52	35.07	28.58
Cryptix_marianmt	40.98	63.36	44.49	35.04	28.55
duth_xanthi_GoogleTranslate_fallback	40.94	62.75	44.21	35.12	28.84
UvA_finetunedMarianMT&finetunedroBERTa	40.85	62.90	44.38	35.03	28.49
Cryptix	40.75	62.54	43.98	34.95	28.69
duth_google_flant5_fallback	40.74	62.60	43.99	34.92	28.65
duth_xanthi_GoogleTranslate	40.73	62.59	43.98	34.91	28.64
duth_xanthi_GoogleTranslate_fallback	40.73	62.60	43.99	34.91	28.63
duth_xanthi_argos	40.49	63.21	44.13	34.73	28.24
pjmathematician_Q25-14	39.08	62.57	43.10	33.19	26.05
pjmathematician_Q25-14	38.49	61.88	42.37	32.62	25.66
pjmathematician_Q25-14	38.24	61.56	42.12	32.40	25.46

BERTScore on Test Data

run ID	P	R	F_1
UvA_finetunedNLLB-1.3B	87.85	87.04	87.42
Skommarkhos_Lucie_SFT	87.74	87.15	87.42
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v4	87.61	87.01	87.28
UvA_finetunedMarianMT	87.72	86.82	87.24
UvA_finetunedNLLB-1.3B&finetunedroBERTa	87.55	86.96	87.23
UvA_finetunedMarianMT&finetunedroBERTa	87.50	86.78	87.11
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v8	87.31	86.91	87.08
duth_xanthi_GoogleTranslate_fallback	87.20	86.77	86.96
duth_xanthi_GoogleTranslate	87.20	86.77	86.96
duth_google_flant5_fallback	87.17	86.74	86.93
Cryptix	87.10	86.63	86.84
Skommarkhos_Lucie_SFT_ARPO	86.79	86.59	86.66
Skommarkhos_Lucie-7B-Instruct-v1.1	86.68	86.46	86.54
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a1	86.68	86.45	86.53
Skommarkhos_Lucie-7B-Instruct-v1.1	86.66	86.46	86.53
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a7	86.67	86.45	86.53
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_arpo_a19	86.67	86.43	86.52
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a5	86.64	86.40	86.49
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a11	86.63	86.39	86.48
pjmathematician_Q25-14	87.00	85.95	86.45
UvA_finetunedT5-base	86.69	86.24	86.44
duth_hybrid_fusion	87.18	85.79	86.43
Cryptix_marianmt	87.17	85.77	86.42
duth_xanthi_argos	87.00	85.91	86.42
yourteamid_marianmt_pun_postedit	87.17	85.74	86.40
Cryptix	87.17	85.74	86.40
duth_xanthi_helsinki	87.17	85.74	86.40
pjmathematician_Q25-14	86.84	85.97	86.37
pjmathematician_Q25-14	86.71	85.89	86.27

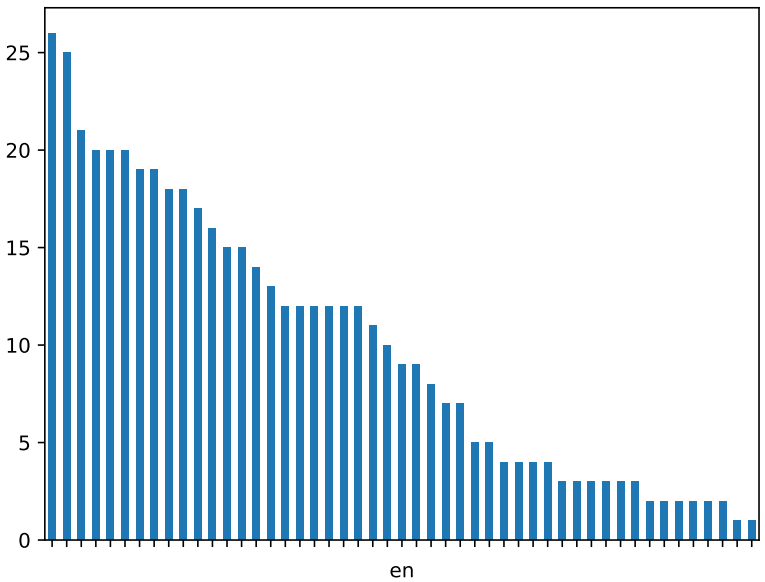
Official Pun Location–based Results

run ID	count	location	%
dsgt_o4_mini_multi_agent_discriminator	1682	156	9.27
dsgt_o4_mini_chain_of_thought_phonetic_embeddings	1682	132	7.85
teamX_aug	1682	118	7.02
pjmathematician_Q25-14	1682	118	7.02
UvA_finetunedMarianMT	1682	114	6.78
UvA_finetunedMarianMT&finetunedroBERTa	1682	114	6.78
Cryptix	1682	113	6.72
Cryptix_marianmt	1682	113	6.72
duth_xanthi_helsinki	1682	113	6.72
yourteamid_marianmt_pun_postedit	1682	113	6.72
duth_xanthi_GoogleTranslate_fallback	1682	112	6.66
duth_hybrid_fusion	1682	112	6.66
Cryptix	1682	112	6.66
duth_google_flant5_fallback	1682	112	6.66
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v4	1682	111	6.60
pjmathematician_Q25-14	1682	111	6.60
duth_xanthi_GoogleTranslate	1682	111	6.60
duth_xanthi_GoogleTranslate_fallback	1682	111	6.60
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v8	1682	111	6.60
UvA_finetunedNLLB-1.3B&finetunedroBERTa	1682	111	6.60
duth_xanthi_argos	1682	109	6.48
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a7	1682	109	6.48
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_arpo_a19	1682	109	6.48
Skommarkhos_Lucie-7B-Instruct-v1.1	1682	109	6.48
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a5	1682	108	6.42
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a11	1682	108	6.42
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a1	1682	108	6.42
Skommarkhos_Lucie-7B-Instruct-v1.1	1682	107	6.36
UvA_finetunedNLLB-1.3B	1682	107	6.36

Manual Evaluation on Test Data

run ID	count	# success	%
dsgt_o4_mini_multi_agent_discriminator	42	37	74
dsgt_o4_mini_chain_of_thought_phonetic_embeddings	42	36	72
duth_xanthi_argos	50	26	52
duth_xanthi_GoogleTranslate_fallback	50	25	50
duth_google_flant5_fallback	50	25	50
Cryptix	50	25	50
duth_xanthi_GoogleTranslate	50	25	50
UvA_finetunedNLLB-1.3B	50	24	48
UvA_finetunedNLLB-1.3B&finetunedroBERTa	50	24	48
duth_xanthi_GoogleTranslate_fallback	48	24	48
pjmathematician_Q25-14	50	23	46
UvA_finetunedT5-base	50	23	46
UvA_T5-base&finetunedroBERTa	50	23	46
Skommarkhos_Lucie_SFT_ARPO	50	22	44
UvA_finetunedMarianMT	50	22	44
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v8	50	22	44
UvA_finetunedMarianMT&finetunedroBERTa	50	22	44
pjmathematician_Q25-14	50	21	42
teamX_aug	50	21	42
Skommarkhos_Lucie_SFT	50	21	42
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v4	50	21	42
teamX_aug	50	20	40
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_arpo_a19	50	20	40
dsgt_simple_mistral_medium	42	20	40
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a1	50	20	40
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a11	50	20	40
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a5	50	20	40
Skommarkhos_Lucie-7B-Instruct-v1.1	50	20	40
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a7	50	20	40

successful translations per EN pun



count	49
mean	9.94
std	7.12
min	1
25%	3
50%	9
75%	15
max	26

The Most Difficult EN Puns

- Without successful translation: "Having too many axe-like tools to do a particular job only adze to the confusion."
- Single successful translation:
  - "The geneticist taught his students how to mendel defective genes" → "*Le généticien a appris à ses étudiants à repriser leurs jeans... et leurs gènes !*" (dsgt_o4_mini_chain_of_thought_phonetic_embeddings)
  - "Volts – the dance you perform after an electric shock" → "*En anglais, le verbe « voltige » désigne une danse après une décharge électrique*" (duth_xanthi_bloomz3b_local)

Frequent Errors (1,297 FR /50 EN)

- 50 cases "[unk]" /3 empty translations /4 incomplete translations
- 72 untranslated texts
- 2 outputs were a mix of EN & FR ("*Horloge en forme de the thinker qui annonce l'heure en disant I think it s 20.25 pm*" and "*I have to keep this fire alight, a crié Tom.*")
- 7 useless repetitions ("*Les vendeurs de chips ne peuvent pas vendre leurs produits, ils ne peuvent pas vendre leurs produits.*")
- 572 translations did not preserve the meaning of the source text
- 80 FR wordplay did not preserve the meaning of the EN pun ("Chips vendors don't get the dough unless their products sell." $\rightarrow$ "*Les agriculteurs ne font pas de blé s'il n'y a pas de blé dans les champs*", or "The scientist had trouble reducing the liquid, he just couldn't concentrate." $\rightarrow$ "*Le brasseur n'arrive pas à maintenir la mousse, pourtant il se fait mousser.*")

BLEU Scores on Training

run ID	score	$n = 1$	$n = 2$	$n = 3$	$n = 4$
Cryptix_rulebased	100.00	100.00	100.00	100.00	100.00
pjmathematician_Q25-14	100.00	100.00	100.00	100.00	100.00
yourteam_rulebased	100.00	100.00	100.00	100.00	100.00
teamX_final	99.77	99.83	99.79	99.76	99.75
UvA_finetunedNLLB-1.3B	54.08	73.99	57.16	48.42	41.77
UvA_finetunedNLLB-1.3B&finetunedroBERTa	53.66	73.57	56.75	48.03	41.34
UvA_finetunedMarianMT	53.38	73.73	56.49	47.55	40.98
UvA_finetunedMarianMT&finetunedroBERTa	52.88	73.28	56.07	47.08	40.43
Cryptix_marianmt	50.09	71.47	52.93	43.97	38.02
Cryptix	50.05	71.52	52.97	44.03	38.08
yourteamid_marianmt_pun_postedit	50.05	71.52	52.97	44.03	38.08
duth_xanthi_helsinki	50.05	71.52	52.97	44.03	38.08
duth_hybrid_fusion	50.02	71.49	52.93	43.98	38.04
Skommarkhos_Lucie_SFT	49.98	72.04	53.64	43.95	36.73
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v4	48.73	71.32	52.46	42.57	35.41
duth_xanthi_argos	47.92	70.75	51.70	42.33	35.73
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v8	47.88	70.80	51.70	41.68	34.45
Skommarkhos_Lucie_SFT_ARPO	46.44	68.89	50.12	40.44	33.32
Skommarkhos_Lucie-7B-Instruct-v1.1	45.92	68.25	49.52	39.93	32.96
Skommarkhos_Lucie-7B-Instruct-v1.1	45.75	68.10	49.35	39.76	32.79
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a1	45.65	68.03	49.25	39.64	32.69
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a7	45.63	68.12	49.23	39.62	32.64
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a11	45.56	68.01	49.18	39.58	32.55
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a5	45.55	67.99	49.17	39.57	32.56
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_arpo_a19	45.50	67.90	49.12	39.50	32.53
UvA_finetunedT5-base	44.93	68.76	48.88	38.65	31.37
UvA_T5-base&finetunedroBERTa	44.22	68.23	48.13	37.94	30.70
Cryptix	43.84	68.61	47.88	37.50	29.99
duth_xanthi_GoogleTranslate_fallback	43.81	68.63	47.86	37.45	29.93

BERTScore on Train

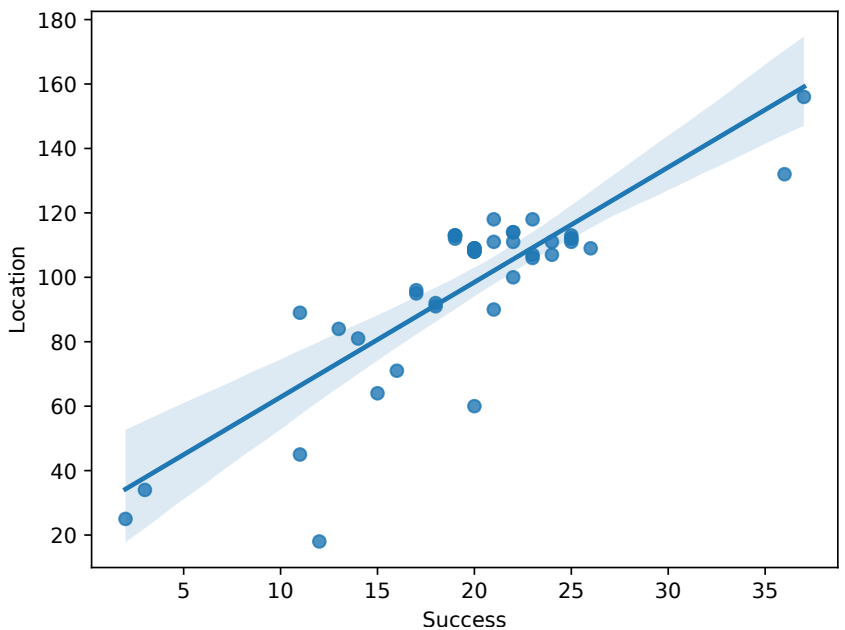
run ID	P	R	F_1
teamX_final	83.81	84.37	84.07
pjmathematician_Q25-14	83.97	83.88	83.90
yourteam_rulebased	83.97	83.88	83.90
Cryptix_rulebased	83.97	83.88	83.90
UvA_finetunedNLLB-1.3B	83.75	82.91	83.30
UvA_finetunedNLLB-1.3B&finetunedroBERTa	83.58	82.89	83.20
UvA_finetunedMarianMT	83.51	82.53	82.99
UvA_finetunedMarianMT&finetunedroBERTa	83.38	82.52	82.92
Skommarkhos_Lucie_SFT	83.32	82.54	82.90
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v4	82.74	82.04	82.36
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v8	82.53	81.97	82.22
UvA_finetunedT5-base	82.28	81.65	81.93
UvA_T5-base&finetunedroBERTa	82.14	81.62	81.85
Cryptix	82.03	81.33	81.65
duth_xanthi_GoogleTranslate	82.02	81.31	81.63
duth_xanthi_GoogleTranslate_fallback	82.02	81.30	81.63
Cryptix_marianmt	82.02	81.17	81.55
yourteamid_marianmt_pun_postedit	82.02	81.16	81.54
duth_xanthi_helsinki	82.02	81.16	81.54
Cryptix	82.02	81.16	81.54
duth_hybrid_fusion	82.02	81.16	81.54
duth_xanthi_GoogleTranslate_fallback	81.74	81.03	81.35
teamX_aug	81.27	81.26	81.17
Skommarkhos_Croissant_SFT	81.34	81.06	81.16
duth_xanthi_argos	81.60	80.73	81.12
Skommarkhos_Croissant_SFT_ARPO	81.27	80.99	81.09
Skommarkhos_Lucie_SFT_ARPO	81.05	80.98	80.97
duth_xanthi_m2m100_1_2B	81.30	80.66	80.94
duth_google_flant5_fallback	81.03	80.42	80.69

Location–based Results on Train

run ID	count	location	%
Cryptix_rulebased	1405	391	27.83
pjmathematician_Q25-14	1405	391	27.83
yourteam_rulebased	1405	391	27.83
teamX_final	1405	380	27.05
dsgt_o4_mini_chain_of_thought_phonetic_embeddings	1405	184	13.10
dsgt_o4_mini_multi_agent_discriminator	1405	183	13.02
UvA_finetunedMarianMT&finetunedroBERTa	1405	178	12.67
UvA_finetunedMarianMT	1405	172	12.24
UvA_finetunedNLLB-1.3B	1405	169	12.03
UvA_finetunedNLLB-1.3B&finetunedroBERTa	1405	169	12.03
teamX_aug	1405	166	11.81
Cryptix_marianmt	1405	162	11.53
duth_hybrid_fusion	1405	161	11.46
Cryptix	1405	161	11.46
yourteamid_marianmt_pun_postedit	1405	161	11.46
duth_xanthi_helsinki	1405	161	11.46
duth_xanthi_argos	1405	154	10.96
UvA_mBARTcc25&finetunedroBERTa	1405	153	10.89
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a5	1405	150	10.68
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a7	1405	149	10.60
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_arpo_a19	1405	149	10.60
Skommarkhos_Lucie-7B-Instruct-v1.1	1405	149	10.60
Skommarkhos_skommarkhos-lucie7binstructv1-1-sft-arpo-a11	1405	148	10.53
Skommarkhos_Lucie-7B-Instruct-v1.1	1405	148	10.53
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v8	1405	146	10.39
Cryptix	1405	145	10.32
UvA_finetunedT5-base	1405	144	10.25
Skommarkhos_skommarkhos_lucie7binstructv1_1_sft_v4	1405	144	10.25
duth_xanthi_GoogleTranslate	1405	143	10.18

Manual Scores VS Location Metrics

- Pearson correlation coefficient between the manual & location-based scores is 0.84
- The highest- and lowest-scoring runs according to the location-based metric align with those identified by the expert evaluation
- Pun location-based evaluation can be a reliable proxy for assessing the quality of wordplay translation



Conclusions on Task 2

- +1,682 new distinct EN source texts with 2,615 FR translations
- 9 teams submitted 52 runs
- Methods: LLMs, commercial MT, out-of-the-box MT, rule-based approaches, & various fine-tuning and training techniques to discriminate wordplay from non-wordplay
- Significant improvements in participants' results compared to the previous years
- BUT the majority of translations do not preserve meaning and wordplay
- Pun location-based evaluation can be a reliable proxy for assessing the quality of wordplay translation

Task 3: Onomastic Wordplay Translation

- *New JOKER 2025 task!*
- Onomastic wordplay is challenging for AI models but common in fictional text (Pokemons, Asterix, Harry Potter,...)
- Parallel corpus of wordplay in named entities in English and French
  - >2,500 FR translations of EN onomastic wordplay
  - Context: a short description of each character/object
  - Different types of wordplay in names: portmanteau, pun/homophone, no manipulation, neologism, assonance/alliteration, anagram...
  - **Train**: 353 onomastic wordplay EN-FR from Asterix & Harry Potter
  - **Test**: 2,333 from video games, literature, and other sources (e.g. Fakemon and alternative or new manual translations–unseen by LLMs)
- Evaluation
  - Matching references
  - Manual evaluation of 1,737 translations of 203 distinct EN wordplay
  - # of untranslated onomastic wordplay

Example


```
[
  {
    "id":"en_1",
    "en":"Asterix",
    "description":"Asterix is the small but
       clever hero of the Asterix comic series.
       Known for his sharp wit and courage, he
       outsmarts the Roman invaders with the help
        of a magical potion that grants him
       superhuman strength. Alongside his loyal
       friend Obelix, Asterix defends his village
        and embodies bravery and cleverness.",
    "fr":"Astérix"
  }
]
```

Official Results for Task 3

run ID	automatic	manual	identical
VerbaNex_gpt4o	39.05	62.56	8.53
pjmathematician_Q332	22.85	46.31	21.82
pjmathematician_Q314	21.13	39.60	33.48
duth_Helsinki	14.83	18.88	77.67
duth_xanthi_Helsinki-NLP-opus-mt-tc-big-en-fr	14.66	18.88	77.45
Cryptix_flanT5	14.49	13.43	38.15
duth_Helsinki	11.83	2.55	100.00
duth_xanthi_facebook-nllb-200-distilled-600M	10.72	16.75	41.83
duth_xanthi_facebook-nllb-200-1.3B	10.72	16.75	41.83
duth_xanthi_MarianMT_BLOOM	10.42	13.86	45.95
duth_xanthi_MarianMT_BLOOM	10.29	13.86	45.78
duth_xanthi_Helsinki-NLP-opus-mt-en-fr	10.29	13.86	45.78
duth_xanthi_t5-base	8.57	7.03	50.32
duth_xanthi_t5-small	8.53	6.00	58.04
duth_xanthi_facebook-m2m100_1.2B	4.71	9.50	19.12
duth_xanthi_facebook-m2m100_418M	4.37	4.00	20.15
team1_gemma2b_v2	4.20	2.99	27.69
duth_hybrid_v1	0.04	1.47	0.21
duth_xanthi_MarianMT_LLM_Prompting	0.00	0.00	0.00
Skommarkhos_Lucie-7B-Instruct_SFT_Q8B_LoRA	0.00	0.00	0.00
copy	11.83	2.55	100.00

Task 3 Results on Train

run ID	automatic	identical
pjmathematician_Q332	55.81	3.40
pjmathematician_Q314	55.81	3.40
duth_xanthi_facebook-m2m100_418M	55.52	3.40
duth_xanthi_Helsinki-NLP-opus-mt-en-fr	55.52	3.40
duth_hybrid_v1	55.52	3.40
duth_xanthi_t5-base	55.52	3.40
duth_xanthi_MarianMT_BLOOM	55.52	3.40
duth_xanthi_Helsinki-NLP-opus-mt-tc-big-en-fr	55.52	3.40
duth_xanthi_MarianMT_LLM_Prompting	55.52	3.40
duth_xanthi_facebook-nllb-200-distilled-600M	55.52	3.40
duth_xanthi_facebook-m2m100_1.2B	55.52	3.40
duth_xanthi_t5-small	55.52	3.40
duth_xanthi_facebook-nllb-200-1.3B	55.52	3.40
duth_Helsinki	55.52	3.40
VerbaNex_gpt4o	11.61	10.48
Cryptix_flanT5	8.50	29.46
duth_Helsinki	5.38	52.69
Skommarkhos_CroissantLLMChat-v0.1_SFT_Q8B_LoRA	4.53	26.35
Skommarkhos_Lucie-7B-Instruct_SFT_Q8B_LoRA	3.97	5.10
team1_gemma2b_v2	1.42	17.28
copy	0.00	0.00

Onomastic Wordplay (Un-)translation

- 172 (10%) were successful
- 12% of reference translations are identical to the EN source wordplay
- The identity baseline ↑ 50% runs
- 50% runs have $>$ 40% FR translations identical to EN
- 30% runs keep $>$ 50% the wordplay instances untranslated
- Among 172 successful translations, 17 were nearly identical to the reference translations – differences manifesting in diacritics, capitalisation, and/or punctuation ("Oreilles de Soie" (run) vs. "Oreilles-De-Soie" (reference) vs. "Ears of Silk" (source))
- 10% of manually evaluated translations (155 instances) were genuinely alternative translations
- untranslated wordplay fails to preserve the intended humorous or pragmatic meaning
- Untranslated onomastic wordplay fails to preserve the intended humorous or pragmatic meaning

Frequent Errors

- 1,737 manually evaluated translations
- 226 generations were identical to EN
- 102 cases with the suffix "-ix" as in Celtic names (overfitting on the names from *Asterix*?)
- 11 cases lack the character's surname
- 29 cases were blank or consisted only of punctuation (e.g., "???")
- 13 cases with spurious overgeneration ("l'aide de") or random translations such as "l'intention des autorités fédérales, il" for "Chimchar"
- 226 cases with extraneous articles (*le, l', la, les*)? e.g."Le Munchlax" for "Munchlax" or "Le Shinx" for "Shinx".
- The Pokémon name "Pidove" was inexplicably translated as "pédophile" in one run

Conclusions on Task 3

- >2,500 wordplay instances in named entities in EN & FR with a short description of each character/object
  - Train: officially translated sources *Asterix* and *Harry Potter*
  - Test: partially new, ↓ overlap with AI training data
- 4 teams submitted 20 runs to Codabench
- Homogeneous results on train: 14 runs with nearly identical scores, high exact matches (56%) and few untranslated names (3.4%)
- Variable results on test
- VerbaNex_gpt4o:
  - 12% exact matches on the training data but 39% on the test
  - stable rate of untranslated names
  - 63% successful translations by manual evaluation (more creative alternative translations?)
- Recurrent errors: untranslated names, overfitting to training data, omission of surnames, and occasional nonsensical generations

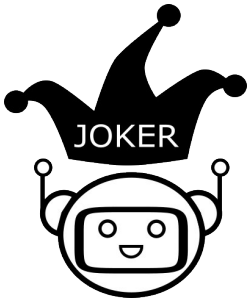
CLEF 2026 JOKER Track

- *Task 1: Humor-aware Information Retrieval: retrieve short humorous texts for a query*
  - *Main IR Task*
  - *Topical relevance models luke-warm, neural ranking degrades...*
  - *But humor-aware IR models improve!*
- *Task 2: Wordplay Translation: translate puns from English to French*
  - *Main NLP Task*
  - *Standard MT models fail: 10% literal translations is a pun*
  - *But humor-aware MT models improve!*
- *Task 3: Onomastic Wordplay Translation*
  - Onomastic (i.e., name-related) wordplay widely used by novelists, comic writers, poets, and playwrights
  - Neologisms – esp. humorous ones – very challenging to translate...
  - Unique corpus of named entities with translations and descriptions
- *JOKER 2022-2024 revisited*: Previous tasks by popular requests (pun detection EN/FR/ES, pun location EN/FR/ES)
- Other unresolved challenges? Future directions?

JOKER Sessions at CLEF 2025

Date	Event
Sep 09 14:15-14:45	JOKER Task Overview Talks
Sep 09 14:45-15:45	*Participant's talks (5x)*
Sep 09 16:30-17:30	*Participant's talks (5x)*
Sep 09 17:30-18:00	Planning Session: Exciting challenges and opportunities, volunteers, roadmap

- Please join the JOKER sessions in Ricardo Marín!
- Program at https://www.joker-project.com/2025/program



Thank you !

Please join our track !

Website : https://joker-project.com[2]

E-mail : contact@joker-project.com

Twitter : https://twitter.com/joker_research

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

[2]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.