Overview of the CLEF 2026 JOKER Task 3: Onomastic Wordplay Translation from English to French

Liana Ermakova[1], Yaël Naud[1], Lilian Binet[1] and Jaap Kamps[2]

[1]HCTI, University of Brest, France
[2]University of Amsterdam, Amsterdam, The Netherlands

Abstract

Onomastic wordplay, i.e., humour based on proper names, remains particularly challenging to translate due to its reliance on linguistic creativity, cultural references, and the interplay between form and meaning. This paper presents an overview of Task 3 of the CLEF 2026 JOKER Lab, which focuses on the automatic translation of English onomastic wordplay into French. We introduce a parallel English-French corpus of onomastic wordplay with descriptions, containing around 3,000 instances from literature, video games, comics, films, and other sources, with many French translations created specifically for this resource. A total of four teams submitted 55 valid runs. The results show that, despite the rapid advances of large language models, translating onomastic wordplay remains a challenging task. Our results show that exact reference matching is insufficient for evaluating creative translation, as manual assessment identified numerous acceptable translations beyond the official references.

Keywords

Onomastic wordplay, Named entities, Machine translation, Humour, Large Language Models, LLM

1. Introduction

Onomastic wordplay, i.e., humour based on proper names, is widely used in literature, comics, films, and video games, but remains one of the most challenging forms of wordplay for both human and machine translation due to its reliance on linguistic creativity, cultural references, and the interplay between form and meaning. This paper presents an overview of Task 3 of the CLEF 2026 JOKER Lab [1], which focuses on the automatic translation of English onomastic wordplay into French. For details on other JOKER-2026 tasks, we refer the reader to their respective overview papers: Task 1 (*Humour-aware Information Retrieval*): Retrieve short humorous texts for a query in English and Hinglish [2], Task 2 (*Pun Translation*): translate puns from English to French [3], and Task 4 (*Humour Generation*): generate humorous text from a given context in English, French, and Spanish [4].

A pilot version of this task on the machine translation of onomastic wordplay was introduced at the JOKER 2022 Lab [5, 6]. It relied on a parallel English–French corpus of onomastic wordplay that we compiled from video games, literature, and other sources. In 2025, we extended the corpus with additional instances and provided short contextual descriptions to facilitate the interpretation and translation of the underlying wordplay [7, 8]. In addition, we transitioned to CodaBench[1] for submission management and evaluation. This year, we continue the 2025 setup and expand the test collection to 2,598 distinct English instances with descriptions and 2,851 French reference translations.

Task 3 involves translating onomastic (i.e., name-based) wordplay from English into French. Onomastic wordplay has been widely used as a rhetorical device by novelists, poets, and playwrights. It is widespread in classic literature, such as in Shakespeare's characters' names, but also names found in modern-day works such as Pokémon, Harry Potter, Astérix, and video games. Meaningful proper names in fictional universes are often neologisms. Neologisms – that is, newly coined words – are among the most common forms of linguistic creativity. Due to their highly idiosyncratic nature, humorous neologisms are challenging for both humans and machines to translate.

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany
liana.ermakova@univ-brest.fr (L. Ermakova); kamps@uva.nl (J. Kamps)
https://www.joker-project.com (L. Ermakova)
0000-0002-7598-7474 (L. Ermakova); 0000-0002-6614-0087 (J. Kamps)

[1]https://www.codabench.org/competitions/8746/

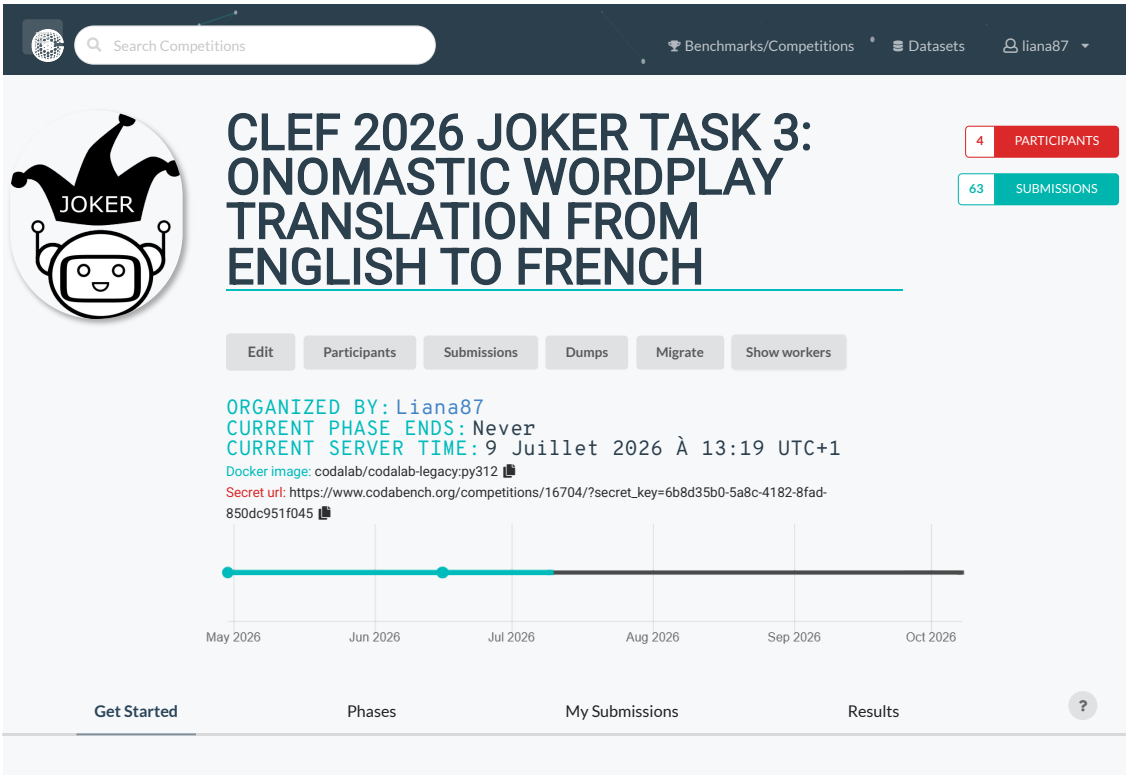

Figure 1: CodaBench for Task 3

A total of 4 teams submitted 63 runs to CodaBench[2] (see Figure 1); among them, 55 runs were valid.

The remainder of this paper is organized as follows. Section 2 describes the training and test collections, their construction, and data format. Section 3 presents the evaluation framework, including the automatic and manual evaluation methodologies. Section 4 summarizes the participating teams and their approaches. Section 5 reports and analyzes the results on both the training and test sets. Finally, Section 6 concludes the paper and discusses future directions.

2. Data

We constructed a parallel corpus of wordplay in named entities in English and French with corresponding descriptions. We collected distinct named entities in English that contain wordplay from video games, movies, literature, and other sources, along with their French translations. The corpus includes named entities from *Astérix, Harry Potter, Pokémon, The Hobbit and The Lord of the Rings, Charlie and the Chocolate Factory, The BFG, Undertale, The Smurfs, Shrek, DC Comics*, and many other sources. The evaluated dataset also includes official and unofficial translations from *the Mortal Engines* and *Fever Crumb* universe, as well as *Fakémon* names. There are 2,598 distinct English instances along with their descriptions with 2,851 French reference translations for testing, and 353 pairs for training. The data includes various types of wordplay, with portmanteaux being the most frequent type. References are produced by translation professionals and students, with the majority being truly novel and not encountered by LLMs during training (for example, our translations of the Fakémon —fan-made Pokémon— names by master's students in translation at the University of Brest).

[2] https://www.codabench.org/competitions/16704/

2.1. Formats

This section outlines the format of the data used in the CLEF 2026 JOKER Task 3.

Input. We provide the training data as JSON files with the following fields:

- **id**: a unique identifier from the input file. Note that this identifier is not unique in the file, as the same English pun might have multiple French translations.
- **en**: the text of the instance of source onomastic wordplay in English. Note that the texts in English are not unique in the file, as the same English pun might have multiple French. translations
- **description**: short contexts for the names and objects, which are often necessary to recognise, understand, and translate the wordplay
- **fr**: translation of the onomastic wordplay into French

For example:

```
[
  {
    "id":"en_89",
    "en":"Winesanspirix",
    "description":"Winesanspirix is a character from the Asterix series,
    ↪  appearing in Asterix at the Olympic Games. He is a Roman wine
    ↪  merchant who sells a particular type of wine. Winesanspirix is not a
    ↪  major character in terms of plot development, but his name and
    ↪  profession add comedic value to the story. His character is part of
    ↪  the satirical approach Asterix takes towards Roman society and its
    ↪  various absurdities.",
    "fr":"Alambix"
  }
]
```

The test data format is identical to that of the training data, except that the field for the target text is omitted.

Output. Participants were asked to submit to CodaBench a ZIP archive containing a file named `prediction.json` in the root directory. This JSON-formatted file was to contain the following fields:

- **run_id**: Run ID starting with `<team_id>_<task_id>_<method_used>` – e.g., `UBO_task_3_gemini-3-flash`
- **manual**: 0 if the run is automatic, or 1 if manual
- **id**: a unique identifier from the input file
- **en**: the text of the instance of source onomastic wordplay in English.
- **fr**: translation of the onomastic wordplay into French

Example:

```
[
  {
    "run_id":"UBO_task3_gemini-3-flash",
    "manual":0,
    "id":"en_89",
    "en":"Winesanspirix",
    "fr":"Vinsanspirix"
  }
]
```

3. Evaluation

Participants' translations are evaluated automatically for accuracy by checking for exact matches against the reference translations. To account for alternative valid translations not found in the reference data, we made use of manual binary evaluation by expert translators. Annotators were provided with the English source named entity, its description, one reference translation for each source named entity, and the submitted translation. Their task was to determine whether the submitted translation was a valid translation of the English named entity. Each translation was evaluated by a single annotator. To reduce the number of manual evaluations, we evaluated only translations that differed from the reference translations and the source English texts. We manually evaluated the translations of 313 English instances of onomastic wordplay from the top three runs of each team, ranked according to the percentage of translations matching the reference translations. For the selected named entities and runs, 21 translations were empty. In total, we manually evaluated 719 distinct French translations. Thus, the official ranking is based on the percentage of translations that match the reference translations or are considered valid according to the human annotation. We report the following scores:

- **Score**: official score over 313 English instances of onomastic wordplay;
- **Refs**: percentage of the translations matching the references over 2,598 English sources;
- **Alter**: percentage of valid alternative but from the references over 313 English sources;
- **Copy**: percentage of non-translations (copy English source into French target) over 2,598 English sources.

4. Participants' approaches

A total of 4 teams submitted 63 submissions; among them, 55 runs were valid.

UBO [9] submitted 5 runs to Task 3. Their approach was based on prompting 4 open models (Qwen3-4B, Mistral-7B, Gemma-2-9b-it, Llama-3.1-8B) and Gemini-3-flash.

University of Amsterdam [10] (team *ezrasmink*) submitted 5 runs based on the fine-tuned MarianMT (Helsinki-NLP/opus-mt-en-fr, 136M parameters), context translation, and copying the source named entity into the target language.

Team *pjmathematician* (no Working Notes submitted) submitted 4 baseline runs.

Team *arampageos* [11] submitted 41 runs to Task 3, combining machine translation with the retrieval strategies.

5. Results

5.1. Official Results

Table 1 provides the official top-5 distinct results per team for Task 3 in terms of the percentage of translations that match the reference translations or are considered to be valid ones according to the human annotation. Besides, we report the percentages of reference matches **Refs**, valid alternative translations **Alter**, and untranslated outputs **Copy**.

The identity run (copying source EN proper name) achieves 13.66% accuracy against references. Many runs, including simple prompting of open LLMs, are outperformed by the non-translated onomastic wordplay. The runs of *arampageos* [11] achieved the best results according to reference matching. Direct prompting of Gemini, even when provided with a description, did not achieve results comparable to those of the *arampageos* [11] team in terms of **Refs**.

The *extraction* run of *ezrasmink* [10] has 98% untranslated English named entities. The best runs of *arampageos* [11] have around 40% of untranslated sources. Among the runs of *UBO* [9], the submission based on Llama has the highest rate of untranslated source (25%). Mistral, Qwen, and Gemma tried to translate almost all named entities. However, their translations were not successful both in terms of matching references and according to manual evaluation.

Table 1
Top-5 results per team for Task 3: Onomastic Wordplay Translation. **Score** is the official score; **Refs, Alter,** and **Copy** denote the percentages of reference matches, valid alternative translations, and untranslated outputs, respectively. **#runs** is the number of runs submitted by each team.

#	team	#runs	Score	Rank	Refs	Alter	Copy	run_id
1	UBO	5	71.88	1	44.11	45.37	8.20	gemini-3-flash
2	arampageos	41	43.13	2	51.58	7.35	40.38	V18_FRLEAD_STRICT_COPY
3	arampageos	41	43.13	3	51.58	7.35	40.38	V18_FRLEAD_WD_STRICT_COPY
4	arampageos	41	43.13	4	51.62	7.35	41.49	V14_OFFICIAL_B04_1700_2299
5	arampageos	41	43.13	5	51.58	7.35	40.38	V21_ASTERIX_DIRECT_ONLY
6	arampageos	41	43.13	6	51.58	7.35	40.38	V21_ASTERIX_COMBO
7	UBO	5	17.25	43	7.08	10.54	24.83	llama
8	ezrasmink	5	13.74	44	6.66	8.31	18.75	marian-ft_context
9	ezrasmink	5	13.74	45	6.97	8.31	18.55	cascade
10	**Baseline**	1	13.10	46	13.66	0.00	100.00	identity_run
11	ezrasmink	5	13.10	47	13.97	0.00	97.69	extraction
12	ezrasmink	5	7.35	49	5.47	2.24	21.21	marian-ft_barename
13	UBO	5	4.47	50	3.54	0.32	4.35	Qwen3-4B
14	pjmathematician	4	2.88	51	4.78	0.00	30.29	baseline_sub
15	UBO	5	2.24	55	2.42	1.60	6.31	mistral
16	UBO	5	1.92	56	1.35	0.64	0.15	gemma

According to the percentage of valid alternative translations (**Alter**), the Gemini run submitted by *UBO* [9] clearly outperformed all other systems, with 45% of its translations judged to be successful alternatives. The second-best result was achieved by Llama (10%), while the best-performing runs of *arampageos* [11] reached only 7.35%.

In terms of official score based on manual evaluation and reference matching over 313 English sources, Gemini is by far the best model with 72% of successful translations. The runs of *arampageos* [11] achieved the second-best score (43%), likely benefiting from their retrieval-based strategy, which also resulted in a high proportion of reference matches. The remaining models achieved substantially lower scores.

In conclusion, state-of-the-art commercial LLMs demonstrate strong performance on onomastic wordplay translation, not only by retrieving official translations but, more importantly, by generating valid alternative translations. Successful genuine translations produced by Gemini accounted for 45% of its 313 translations of English sources. Overall, Gemini generated genuine translations for 235 of the 313 English source instances, of which 142 were judged successful and 93 unsuccessful. Despite this remarkably high and somewhat unexpected performance, the model still failed in approximately 40% of the cases requiring a genuine translation, indicating that onomastic wordplay translation remains far from a solved problem.

5.2. Train results

Table 2 presents the top-5 runs per team in terms of reference matching on the training set. The retrieval-based runs of *arampageos* [11] achieved nearly perfect performance (99%) by reproducing the reference translations from the *Astérix* and *Harry Potter* datasets. In contrast, the Gemini run submitted by *UBO* [9] achieved only 40% reference matches. Note, Gemini was not explicitly prompted to retrieve existing translations from the web. Several systems simply copied the English named entities, while *marian-ft_barename* and the *UBO* runs based on Llama, Gemma, Qwen, and Mistral attempted to generate translations but achieved little success.

Table 2
Train data top-5 results per team for Task 3: Onomastic Wordplay Translation. **Refs** and **Copy** denote the percentages of reference matches and untranslated outputs, respectively. **#runs** is the number of runs submitted by each team.

#	team	#runs	Rank	Refs	Copy	run_id
1	arampageos	41	1	98.58	10.48	SAFE_NOPAREN
2	arampageos	41	2	98.58	10.48	POKEMON_FULL
3	arampageos	41	3	98.58	10.48	POKEMON_SAFE
4	arampageos	41	4	98.58	10.48	LEXRULES
5	arampageos	41	5	98.02	10.48	V9_SANITIZE_BIG
6	UBO	5	42	39.66	10.20	gemini-3-flash
7	ezrasmink	5	43	13.03	97.17	extraction
8	ezrasmink	5	45	11.05	17.56	cascade
9	pjmathematician	4	46	10.48	100.00	baseline_sub
10	ezrasmink	5	47	10.48	17.56	marian-ft_context
11	Baseline	1	50	10.48	100.00	indentity_run
12	ezrasmink	5	52	9.63	25.78	marian-ft_barename
13	UBO	5	53	5.67	25.50	llama
14	UBO	5	54	3.12	11.33	mistral
15	UBO	5	55	2.55	0.00	gemma
16	UBO	5	56	2.27	2.55	Qwen3-4B

Table 3
Distribution of translation problems identified during the manual analysis, with illustrative examples.

Problem type	#	Example (EN)	Example (FR)
Random	77	*Boltling*	*Bounteau*
Partially random	66	*Orchatter*	*Orchijace*
Untranslated	18	*Trogglehumper*	*Troglehumper*
Partially untranslated	140	*Marmadoc 'Masterful' Brandybuck*	*Marmadoc 'Masterful' Brandebouc*
Literal translation	20	*Bonnie Hopps*	*Bonnie Gambade*
Partially literal translation	9	*Moro Burrows*	*Moro Creusot*
Lost wordplay/reference	108	*Nightmare Moon*	*La Lune de Nuit*
Incorrect grammar/spelling	7	*Savoir Fare*	*Savis Fare*
Incomplete or truncated	46	*Emotimara*	*Em*
Empty	22		
Other	106	*Barrowdowns*	*Ardenne*
None	189	*Ash Mountains*	*Monts de Cendres*

5.3. Manual Analysis

Among the 740 evaluated translations, 22 were **empty**. Overall, 527 translations (71%) were judged **invalid**. The detailed statistics of error types in the evaluated dataset is given in Table 3. Note, that for some cases annotators marked more than one error type.

The most frequent error was **partial non-translation**, where only a small part of the name was adapted while the rest remained untranslated and semantically opaque to French readers, observed in 140 instances. For example, the English *"Marmadoc 'Masterful' Brandybuck"* was translated as *"Marmadoc 'Masterful' Brandebouc,"* leaving the nickname untranslated. Such partial translations hinder comprehension for French readers and eliminate the intended humorous effect. This behavior may have several explanations: the model may assume that proper names should not be translated; it may consider part of the English wordplay understandable to a French audience; or it may be unable to generate an appropriate wordplay and therefore retain the original English expression. In this example, the second explanation is the most likely, as Masterful appears to be interpreted as a title rather than a nickname. The official French translation, however, correctly renders it as *"Marmadoc 'Maindemaître'*

Brandibouc."

We also observed a related phenomenon in a small number of cases, where the English wordplay was neither translated nor replaced but only minimally adapted to French through minor orthographic or phonetic modifications. We refer to this strategy as **Frenchization**, whereby the original wordplay is retained while being slightly altered to better conform to French spelling or pronunciation. Examples include *Mareep* becoming *Marieep* and *Mirkwood* becoming *Mirkwoode*. The Pokémon *Geodude* (from geo and dude) was translated as *Géodude*. Although Géo is understandable in French, dude remains untranslated and is unlikely to be understood by a French-speaking audience. A similar phenomenon can be observed with the Fakémon (fan-made Pokémon) *Locustalk* (from "locust", species of grasshoppers, and "stalk", a part of plants; they are big grasshoppers with stalk-like antennas), translated as *Locustalque*. Rather than translating the name, the model merely altered the spelling of the suffix to make it appear more French, while preserving a form that carries no meaning in French.

Another frequent issue was the **loss of the original wordplay and/or its references**, observed in 108 evaluated translations. This is unsurprising, as machine translation has long struggled with humor, which often relies on subtext, ambiguity, double meanings, phonetic similarity, and context-specific references that are rarely preserved in non-human translation. For example, *Nightmare Moon* was translated as *"La Lune de Nuit,"* resulting in the loss of both the double meaning of *Nightmare*, which can refer to both a bad dream and a night mare (a female horse), as an animal is lost, and the broader reference to mares. Another example is *Gorgun*, a blend of "gorgon" and "gun", referring to a half-human, half-snake creature carrying a submachine gun, which was translated as *Gorgogne*. While the reference to the gorgon ("gorgone" in French) is preserved, the wordplay involving gun is lost. An even more striking example is the small panda-looking Pokémon *Pancham*, whose name combines panda, punch, and champion. It was translated as *Klangou*, a name that may instead evoke kangaroo ("kangourou" in French), leaving French readers wondering why a panda is named after a kangaroo.

In the evaluated set, we found 20 **literal translations** such as *Bonnie Hopps / Bonnie Gambade* and 9 **partially literal translations**, e.g., *Moro Burrows / Moro Creusot*.

A total of 77 translations were classified as **random**, as they bore no identifiable semantic or lexical relationship to the source entity (e.g., *Boltling* / *Bounteau*). An additional 66 translations were judged **partially random**, where only part of the source name appeared to be preserved while the remaining elements were replaced arbitrarily (e.g., *Orchatter* / *Orchijace*).

A total of 18 instances were **not translated** but had small spelling variations, e.g., *Trogglehumper / Troglehumper*.

A total of 46 translations were **truncated or incomplete**.

Nevertheless, some translations 213 (29%) **successfully** preserve the original wordplay or its underlying references while remaining natural for a French-speaking audience. Among them, 189 were identified as having no issues, while 24 exhibited only minor issues and were still considered to be valid translations. For example, a Fakémon *Projectoad* (from "projectile" and "toad"), a toad that fires darts, is translated as *Crapojectile*. This is an alternative translation to our reference *Sarbatracane* (from "sarbacane", blow gun, and "batracien", a species of amphibians). The translation preserves both references: toad ("crapaud" in French) and projectile. Another successful example is *Hitmongyro*, whose name combines "hit", "monster", and "gyro", referring to a Fakémon that spins on its head while kicking. It is a Fakémon whose name begins with "hitmon," a prefix shared with some other official Pokémons (Hitmonchan and Hitmonlee, for example). The model translates it as *Tourbignon*, a blend of "tourbillon" ("whirlwind") and "gnon" ("thump"). This is a particularly effective adaptation, as "gnon" also echoes the French Pokémon name *Tygnon* (English *Hitmonchan*), preserving both the intertextual Pokémon reference and the idea of rotation.

6. Discussion and Conclusions

In this paper, we have described the CLEF 2026 JOKER Task 3: Onomastic Wordplay Translation. We constructed a parallel English-French corpus of onomastic wordplay with descriptions, collected

from literature, video games, comics, films, and other sources, with a significant portion of our own translations. The dataset comprises 2,598 English named entities, 2,851 French reference translations for testing, and 353 training pairs. A total of 4 teams submitted 55 valid runs.

Our main findings are the following. First, the best-performing systems, according to reference matching, were retrieval-based runs (e.g., *arampageos*), which correctly reproduced more than half of the reference translations. However, these systems struggle to produce genuine onomastic wordplay translations. Second, the copy baseline outperformed several systems, including direct prompting of Qwen, Mistral, and Gemma. Third, Gemini demonstrated that modern frontier LLMs can successfully produce valid translations in a substantial proportion of cases. Successful genuine translations produced by Gemini accounted for 45% of its 313 translations of English source names. Even when provided with a description explaining the meaning of the named entity, Gemini frequently generated new names instead of retrieving existing French translations. Importantly, Gemini was not explicitly prompted to retrieve existing translations from the web, suggesting that retrieval or retrieval-augmented prompting could further improve performance.

Overall, the manual analysis of 740 genuine translations, different from our references, showed that only 29% were successful. The error analysis revealed that the most frequent problems were: partially untranslated names (19%); loss of the original wordplay or reference (15%); random or partially random inventions (19%); incomplete, empty, or truncated translation (9%); literal translations that sacrifice the joke; omission of parts of compound names; missing punctuation or minor grammatical issues.

Automatic evaluation based solely on reference matching underestimates translation quality. Manual evaluation identified numerous valid translations that differed from the official reference. Consequently, exact reference matching alone is insufficient for evaluating creative translation tasks, and human assessment remains necessary to recognize acceptable alternative solutions.

Despite the remarkable success of Gemini, the model still failed in approximately 40% of the cases requiring a genuine translation, demonstrating that the translation of onomastic wordplay remains far from solved.

For more information about the JOKER lab this year and other tasks, please refer to the overview paper [1] and the Working Notes papers for Task 1: Humour-aware Information Retrieval [2], Task 2: Pun Translation [3], and Task 4: Humour Generation [4]. Visit the JOKER website at https://joker-project.com for any other information related to the track.

Acknowledgments

We thank the Master's students in translation and technical writing from the University of Brest for participating in data annotation, and CodaBench for hosting the competition.

This project is funded by the National Research Agency (ANR-19-GURE-0001) under the program "*Investissements d'avenir*" integrated into France 2030. Jaap Kamps is partly funded by the Netherlands Organization for Scientific Research (NWO NWA #1518.22.105), the University of Amsterdam (AI4FinTech program), and ICAI (AI for Open Government Lab). Views expressed in this paper are not necessarily shared or endorsed by those funding the research.

Declaration on Generative AI

During the preparation of this work, the authors used *ChatGPT* and *Grammarly* in order to: **Grammar and spelling check** and **Paraphrase and reword**. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] L. Ermakova, et al., Overview of the CLEF 2026 JOKER track: Humor detection, search, and translation, in: M. Hagen, M. Potthast, B. Stein, P. Schaer, E. Zangerle, S. MacAvaney, J. M. Struß,

E. Sánchez Salido, A. Barrón Cedeño, A. García Seco de Herrera (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Seventeenth International Conference of the CLEF Association (CLEF 2026), Lecture Notes in Computer Science, Springer, 2026.

[2] P. Vachharajani, A. Mukherjee, J. Singh, K. Tewari, S. Pal, J. Kamps, L. Ermakova, Overview of the CLEF 2026 JOKER Task 1: Humor-Aware Information Retrieval in English and Hinglish, in: [12], 2026.

[3] L. Ermakova, Y. Naud, L. Binet, B. Dumas, J. Kamps, Overview of the CLEF 2026 JOKER Task 2: Pun Translation from English to French, in: [12], 2026.

[4] I. Kuzmin, A.-G. Bosser, J. Kamps, L. Ermakova, Overview of the CLEF 2026 JOKER Task 4: Humor Generation, in: [12], 2026.

[5] L. Ermakova, T. Miller, F. Regattin, A. Bosser, C. Borg, É. Mathurin, G. L. Corre, S. Araújo, R. Hannachi, J. Boccou, A. Digue, A. Damoy, B. Jeanjean, Overview of joker@clef 2022: Automatic wordplay and humour translation workshop, in: A. Barrón-Cedeño, G. D. S. Martino, M. D. Esposti, F. Sebastiani, C. Macdonald, G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction - 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy, September 5-8, 2022, Proceedings, volume 13390 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 447–469. URL: https://doi.org/10.1007/978-3-031-13643-6_27. doi:10.1007/978-3-031-13643-6_27.

[6] L. Ermakova, T. Miller, J. Boccou, A. Digue, A. Damoy, P. Campen, Overview of the CLEF 2022 JOKER task 2: Translate wordplay in named entities, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 1666–1680. URL: https://ceur-ws.org/Vol-3180/paper-127.pdf.

[7] L. Ermakova, R. Campos, A. Bosser, T. Miller, Overview of the CLEF 2025 JOKER lab: Humour in machine, in: J. Carrillo-de-Albornoz, A. G. S. de Herrera, J. Gonzalo, L. Plaza, J. Mothe, F. Piroi, P. Rosso, D. Spina, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction - 16th International Conference of the CLEF Association, CLEF 2025, Madrid, Spain, September 9-12, 2025, Proceedings, volume 16089 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 315–337. URL: https://doi.org/10.1007/978-3-032-04354-2_18. doi:10.1007/978-3-032-04354-2_18.

[8] L. Ermakova, T. Miller, Y. Naud, A. Bosser, R. Campos, Overview of the CLEF 2025 JOKER task 3: Onomastic wordplay translation, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2025, Madrid, Spain, 9-12 September 2025, volume 4038 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 2780–2790. URL: https://ceur-ws.org/Vol-4038/paper_220.pdf.

[9] L. Ermakova, Y. Naud, L. Binet, B. Dumas, Testing LLMs on JOKER Task 3: Onomastic Wordplay Translation and Task 4: Humour Generation. UBO at CLEF 2026, in: [12], 2026.

[10] J. van Bakel, R. Flier, E. Smink, J. Bakker, J. Kamps, University of Amsterdam at the CLEF 2026 JOKER Track, in: [12], 2026.

[11] G. Arampatzis, A. Arampatzis, DUTH at CLEF JOKER 2026: Conservative Hybridisation for Humor-Aware Retrieval, Pun Translation, Onomastic Wordplay Translation, and Multilingual Humour Generation, in: [12], 2026.

[12] E. Sánchez Salido, A. Barrón Cedeño, A. García Seco de Herrera, S. MacAvaney, J. M. Struß (Eds.), Working Notes of CLEF 2026: Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, CEUR-WS.org, 2026.