

# Overview of the CLEF 2022 JOKER Task 1: Classify and Explain Instances of Wordplay

Liana Ermakova<sup>1,2</sup>, Fabio Regattin<sup>3</sup>, Tristan Miller<sup>4</sup>, Anne-Gwenn Bosser<sup>5</sup>,  
Sílvia Araújo<sup>6</sup>, Claudine Borg<sup>7</sup>, Gaelle Le Corre<sup>8</sup>, Julien Boccou<sup>1</sup>,  
Albin Digue<sup>1</sup>, Aurianne Damoy<sup>1</sup>, Paul Campen<sup>1</sup> and Orlane Puchalski<sup>1</sup>

<sup>1</sup>Université de Bretagne Occidentale, HCTI, 29200 Brest, France

<sup>2</sup>Maison des sciences de l'homme en Bretagne, 35043 Rennes, France

<sup>3</sup>Dipartimento DILL, Università degli Studi di Udine, 33100 Udine, Italy

<sup>4</sup>Austrian Research Institute for Artificial Intelligence, Vienna, Austria

<sup>5</sup>École Nationale d'Ingénieurs de Brest, Lab-STICC CNRS UMR 6285, France

<sup>6</sup>Universidade do Minho, CEHUM, 4710-057 Braga, Portugal

<sup>7</sup>University of Malta, Msida MSD 2020, Malta

<sup>8</sup>Université de Bretagne Occidentale, CRBC, 29200 Brest, France

<sup>8</sup>Université de Bretagne Sud, HCTI, 56321 Lorient, France

## Abstract

As a multidisciplinary field of study, humour remains one of the most difficult aspects of intercultural communication. Understanding humor often involves understanding implicit cultural references and/or double meanings, which raises the questions of how to detect and classify instances of this complex phenomenon. This paper provides an overview of Pilot Task 1 of the CLEF 2022 JOKER track, where participants had to classify and explain instances of wordplay. We introduce a new classification of wordplay and a new annotation scheme for wordplay interpretation suitable both for phrase-based wordplay and wordplay in named entities. We describe the collection of our data, our task setup, and the evaluation procedure, and we give a brief overview of the participating teams' approaches and results.

## Keywords

wordplay, computation humour, pun, classification, wordplay interpretation, wordplay location, deep learning

## 1. Introduction

Creative language, such as humour and wordplay, is all around us: from entertainment to advertisements to business relationships. Internet humour flourishes on social networks, special humour-dedicated websites, and on web pages focusing on edutainment or infotainment [1]. As a multidisciplinary research area, humour has

---

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

✉ liana.ermakova@univ-brest.fr (L. Ermakova)

🌐 <https://www.joker-project.com/> (L. Ermakova)

🆔 0000-0002-7598-7474 (L. Ermakova); 0000-0003-3000-3360 (F. Regattin); 0000-0002-0749-1100

(T. Miller); 0000-0003-4321-4511 (S. Araújo); 0000-0003-3858-5502 (C. Borg)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

been a focus of interest to many academics from different theoretical backgrounds. Many humour-related academic studies have been centred on culture-specific humour [2, 3] and the (un)translatability of this universal phenomenon [4, 5], which poses genuine challenges for subtitlers and other translators. In order to better understand and process this creative use of language, we have to recognise that it requires special treatment, not only insofar as linguistic mechanisms are concerned, but also regarding the universe of paralinguistic elements [6]. From a metalinguistic/metadiscursive point of view, wordplay includes a wide variety of dimensions that exploit or subvert the phonological, orthographic, morphological, and semantic conventions of a language [7, 8]. It is therefore vitally important that natural language processing applications be capable of recognising and appropriately dealing with instances of wordplay. And indeed, numerous studies have already been conducted for the related task of humour generation [9, 10, 11, 12, 13, 14].

To make a step forward to the automation of humour and wordplay analysis, we introduced the JOKER track at CLEF 2022. The goal is to bring together translators and computer scientists to further the computational analysis of humour. Our workshop proposed three pilot tasks [15]:

**Pilot Task 1:** classify and explain instances of wordplay,

**Pilot Task 2:** translate wordplay in named entities, and

**Pilot Task 3:** translate entire phrases containing wordplay (puns).

This paper covers Pilot Task 1. We introduce a new classification of wordplay and we discuss the shortcomings of previous classifications from the literature. Moreover, we propose a new annotation scheme for wordplay interpretation. Our interpretation scheme is applicable both for phrase-based wordplay, including puns, and wordplay within named entities, including portmanteaux. We present an evaluation benchmark of our own devising and we present and discuss the participating systems and their results.

The paper is organised as follows. In §2 we provide the definitions of wordplay from the literature as well as existing classifications. In §3 we describe the initial dataset created in the JOKER project. The data was annotated, at first, according to the classifications from the literature. However, as we noticed that its classes overlapped, we introduced our own classification system and used it for our final annotation, described in §4. The corpus details can be found in §5. Results from our own preliminary experiments on wordplay perception are found in §6, and alternative classification methods proposed by our participants are covered in §7. The methods, evaluation metrics, and results are described in §§8, 9, and 10, respectively. Some concluding remarks are given in §11.

## 2. Related work

In this section, we first define the concepts of “wordplay” and “pun” and then present different proposed classifications for these complex concepts.

### 2.1. Definitions

The definitions given by mass-market dictionaries are very close:

**Definition 2.1.** Wordplay is the activity of joking about the meaning of words, especially in a clever way. (Cambridge Learner’s dictionary)

**Definition 2.2.** Wordplay is the clever or amusing use of words, especially involving a word that has two meanings or different words that sound the same. (Oxford Advanced Learner’s dictionary)

**Definition 2.3.** Pun is the clever or humorous use of a word that has more than one meaning, or of words that have different meanings but sound the same. (Oxford Advanced Learner’s dictionary)

**Definition 2.4.** Pun is a humorous use of a word or phrase that has several meanings or that sounds like another word. (Cambridge Learner’s dictionary)

Delabastita [16] and Gottlieb [17] consider these terms synonymous and generally interchangeable. According to Delabastita [16], a wordplay or pun

is the general name for the various textual phenomena in which structural features of the language(s) used are exploited in order to bring about a communicatively significant confrontation of two (or more) linguistic structures with more or less similar forms and more or less different meanings.

Chiaro [18] and Giorgadze [19] make a distinction between the terms “wordplay” and “pun”. According to Giorgadze [19], “wordplay can be discussed in its narrow and broad senses”. In the broad sense, the term is a hypernym which may subsume (without limitation) the following phenomena:

**puns** e.g., “Just as a poached egg isn’t a poached egg unless it’s been stolen from the woods in the dead of the night.” (Roald Dahl, *Charlie and the Chocolate Factory*)

**wellerisms** e.g., “‘Don’t move, I’ve got you covered,’ said the wallpaper to the wall.”; “‘We’ll have to rehearse that,’ said the undertaker, as the coffin fell out of the car.”

**spoonerisms** e.g., “Time wounds all heels” instead of “Time heals all wounds”; “a well-boiled icicle” instead of “a well-oiled bicycle”

**anagrams** e.g., “genuine class” for “Alec Guinness” (Dick Cavett)

**palindromes** e.g., “No lemons, no melon!”; “Straw? No, too stupid a fad! I put soot on warts.”

**onomatopoeia** e.g., “Water plops into pond, splish-splash downhill warbling magpies in tree trilling, melodic thrill. . .” (Lee Emmet, “Running Water”)

**mondegreens** e.g., “Lady Mondegreen” instead of “layd him on the green”

**malapropisms** e.g., “the very pineapple of politeness”; “Medieval victims of the Bluebonnet plague”

**neologisms and portmanteaux** e.g., “meringued: the act of being changed into, or being trapped inside, a large meringue” (Jasper Fforde, *The Eyre Affair*); “ChameleoCar: a multi-coloured car that can change hue at the flick of a switch” (ibid.); “frumious” from “fuming” and “furious” (Lewis Carroll, “The Hunting of the Snark”)

**alliteration, assonance, and consonance** e.g. “Peter Piper picked a peck of pickled peppers”

In a narrower sense, a wordplay is a pun when it is based on the ambiguity that occurs when a word or phrase has more than one meaning and can be understood in different ways.

Giorgadze [19] claims that the terms “pun” and “wordplay” are synonymous when the latter is understood in its narrow sense. Redfern [20] considers that “to pun is to treat homonyms as synonyms.” This ambiguity can be based on

- lexis and semantics, the lexical ambiguity of a word or phrase pertaining to its having more than one meaning in the language to which the word belongs; or
- syntax, where a sentence may have two (or more) different interpretations according to how it is parsed.

Delabastita [16] argues that puns are textual phenomena, meaning that they are dependent on the structural characteristics of language as an abstract system. The context should thus be taken into account when analysing puns. The context may be situational or textual. As Gottlieb [21, p. 210] writes,

The intended effect of wordplay can accordingly be conveyed through dialogue (including intonation and other prosodic features), combined with non-verbal visual information, or through written text. . .

## 2.2. Classification of Wordplay

Delabastita [16] distinguishes the following categories:

**Phonological and graphological structure** This is wordplay based on sound or spelling – e.g., “Love at first bite.” This category includes:

**Homonymy** (identical sounds and spelling)

**Homophony** (identical sounds but different spellings)

**Homography** (different sounds but identical spelling)

**Paronymy** (there are slight differences in both spelling and sound)

**Lexical structure (polysemy and idioms)** This concerns the distance between the idiomatic and literal reading of idioms – e.g., “Britain going metric: give them an inch and they’ll take our mile.”

**Morphological structure** This type of wordplay is based on the distinction between the accepted meaning of the words and the interpretation of the components – e.g., “‘I can’t find the oranges,’ said Tom fruitlessly.”

**Syntactic structure** Here, grammar generates puns through sentences or phrases that can be understood in more than one way. For example, “How do you stop a fish from smelling? Cut off its nose.” [19, p. 274]

Delabastita [16] also distinguishes *horizontal* from *vertical* puns. In horizontal puns, the secondary meaning is expressed concretely within the text: “The mere nearness of the pun components may suffice to bring about the semantic confrontation; in addition, grammatical and other devices are used to highlight the pun.” [16, p. 129] In vertical wordplay, the pivotal element is mentioned only once: “one of the pun’s components is materially absent from the text and has to be triggered into semantic action by contextual constraints.” [16, p. 129] Some examples are given in Figure 1.

Gottlieb [17] considers wordplays and puns as synonymous linguistic units and divides them up into three categories:

**Lexical homonymy** The central feature is single-word ambiguity.

**Collocational homonymy** The word-in-context ambiguity is the central feature.

**Phrasal homonymy** The clause ambiguity is the central feature.

Giorgadze [19], on the other hand, breaks the term pun into three other categories:

**Lexical-Semantic Pun** (homonyms, homophone, polysemantic words) – e.g., “I like kids, but I don’t think I could eat a whole one.” (the polysemous word “kid” creating the pun); “Where do fish learn to swim? They learn from a school.” (Lewis Carroll, *Alice’s Adventures in Wonderland*)

**Structural-Syntactic Pun** Here, a complex phrase or sentence can be understood in different ways. For example, “‘I rushed out and killed a huge lion in my pajamas.’ ‘How did the lion get in your pajamas?’”

|            | vertical   | horizontal  |
|------------|--|---|
| homonymy   | There was a brass plate screwed on the wall beside the door. It <b>said</b> : "C V Cheesewaller, DM (Unseen) B. Thau, BF." It was the first time Susan had ever heard metal speak. (T. Pratchett, <i>Soul Music</i> , quoted in [22, p. 29]) | Well, yah, dey lose <b>members</b> in there. Their <b>members</b> lose <b>members</b> (T. Pratchett, <i>Soul Music</i> , quoted in [22, p. 19])   |
| homophony  | <b>Beleave</b> in Britain. ( <i>The Sun</i> )  | Why can a man never starve in the Great Desert? Because he can eat the <b>sand which is</b> there. But what brought the <b>sandwiches</b> there? Why, Noah sent <b>Ham</b> , and his descendants <b>mustered</b> and <b>bred</b> . How the <b>US</b> put <b>US</b> to shame. [23]                               |
| homography | You can tune a guitar, but you can't tuna fish. Unless you play <b>bass</b> .  |   |
| paronymy   | <b>Landen Parke-Laine</b> (i.e., "land on Park Lane", referring to the Monopoly board game) (Jasper Fforde, <i>Thursday Next</i> series)   | "That's a <b>bodacious</b> audience," said Jimbo.<br>"Yeah, that's right, <b>bodacious</b> ," said Scum. "Er. What's bodacious mean?"<br>"Means... means it <b>bodes</b> ," said Jimbo.<br>"Right. It looks like it's <b>boding</b> all right."<br>(Terry Pratchett, <i>Soul Music</i> , quoted in [22, p. 19]) |

**Figure 1:** Examples of horizontal and vertical puns.

**Structural-Semantic Pun** Per Giorgadze [19, p. 274], "Structural-semantic ambiguity arises when a word or concept has an inherently diffuse meaning based on its widespread or informal usage." This is mainly found with idiomatic expressions: " 'Did you **take a bath**? ' 'No, only towels, is there one missing?' " [19, p. 275]

Chuandao [24] believes that a pun cannot be reduced to play on the meaning and the homophony of a word and considers that the context, the logic and the way the pun is formulated should also be taken into account. Chuandao defines the following categories:

**Homonymic pun** (identical sounds and spelling)

**Lexical meaning pun** (polysemous words)

**Understanding pun** (absence of any pun word, but the context enables the addressee to understand the implied meaning of a sentence) – for example:

My sister Mrs. Joe Gargery, was more than twenty years older than I, and had established a great reputation with herself and the neighbours because she **had brought me up "by hand"**. Having at that time to find out for myself what the expression meant, and knowing her to have **a hard and heavy hand**, and to be much in **the habit of laying it upon her husband as well as upon me**, I supposed that Joe Gargery and I were both **brought up by hand**." (Charles Dickens, *Great Expectations*, quoted in [24])

**Figurative pun** (opposition between the surface and figurative meaning of a simile or metaphor) – examples:

In reply, Dr. Zunin would claim that a little practice can help us feel comfortable about changing our social habits. We can become accustomed to any changes we choose to make in our personality. "It's like getting used to a new car. It may be unfamiliar at first, but it goes much better than the old one." [24]

Here, the simile "becoming accustomed to any changes in our social habits" is compared to "getting used to a new car".

**Logic pun** (a kind of implication in a given context) –for example:

*Lady Capulet:* . . . Some grief shows much of love;  
But much of grief shows still some want of wit.

*Juliet:* Yet let me weep for such **a feeling loss**.

*Lady Capulet:* So shall you **feel the loss**, but not the friend.  
Which you weep for.

*Juliet:* Feeling **so the loss**.  
I cannot choose but ever weep the friend.

*Lady Capulet:* Well, girl, thou weep'st not so much for this death,  
As that the villain lives which slaughter'd him.

*Juliet:* What villain, madam?

*Lady Capulet:* That same villain, Romeo.

*Juliet:* Villain and he are many miles asunder. —  
God pardon him! I do with all my heart;  
And yet no man like he doth grieve my heart.

*Lady Capulet:* That is, because the traitor murderer lives.

*Juliet:* Ay, madam, from the reach of these my hands.  
Would none but I might venge my cousin's death!

*Lady Capulet:* We will have vengeance for it, fear thou not:  
Then weep no more. I'll send to one in Mantua,  
Where that same banish'd runagate doth live,

Shall give him such an unaccustom'd dram,  
That he shall soon keep Tybalt company:  
An then, I hope, thou wilt be satisfied.

*Juliet*: Indeed, I never shall be satisfied  
With Romeo, till I behold him — dead —  
(William Shakespeare, *Romeo and Juliet*)

In this excerpt, Lady Capulet uses the phrase “feel the loss” to refer to Juliet’s grief for losing her cousin, while Juliet goes on and expresses her grief for losing Romeo.

### 3. Initial data annotation

At the very beginning of the project, we collected more than 1000 translated wordplay instances in English and French from various sources: video games (*Enter the Gungeon*, *Undertale*, *South Park*, *League of Legends*, *Phoenix Wright*, *Pokémon*, etc.), advertising slogans, literature (Shakespeare, *Alice’s Adventures in Wonderland*, *Asterix*, *How to Train your Dragon*, *Harry Potter*, etc.). Some slogans and punning tweets were translated by our experts.

As previously mentioned, in its broad sense, a wordplay includes sub-categories such as puns, wellerisms, spoonerisms, anagrams, palindromes, onomatopoeia, mondegreens, malapropisms, neologisms and portmanteaux, alliterations, assonance, and consonance. In its narrow sense, a wordplay is a pun based on the ambiguity that occurs when a word or phrase has more than one meaning and can be understood in different ways.

Our research project considers wordplay in its broad sense. Following Delabatista’s classification, our corpus mainly includes two types of wordplay annotation based on sound and spelling (phonological and graphological structure) and lexical structure (polysemy and idioms). Only a few puns play on syntactic and morphological structures.

The collected data mainly contains punning named entities, in many cases neologisms. Each pun in each language was classified in several classes according to a well-defined multi-label classification and explained with respect to how the pun is constructed. For example, the punning joke “Why is music so painful? Because it hertz” was annotated as “Paronymy” under the Structure classification (there being slight differences in both spelling and sound) and explained simply as “hertz/hurts”.

The total amount of English wordplay instances, after being classified according to Type (type of humour) and/or Structure (wordplay based on phonological and graphological structure), are given by frequency in Tables 1 and 2. For each category, we provide an example and its translation.

The “others” category in Table 1 refers to anagrams, mondegreens, onomatopoeia, spoonerisms, and wellerisms. Some of the collected entries (342 entries in the English corpus) employed a type of humour not directly related to wordplay (e.g.,



**Table 1**

Statistics and examples for the initial classification of wordplay type.

| Type                                    | # entries | English  | French  |
|---|-----------|--|---|
| Neologisms and portmanteaux             | 1353      | Cat lovers will only drink their kit-TEA. (Lipton ad)  | Les amoureux des chats ne boivent que du thé Mat-chat.                              |
| Puns                                    | 586       | Hellmann’s makes chicken so juicy, all the competition is squawking ( <i>Hellmann ad</i> )         | Avec Hellmann’s, le poulet est si juteux que la concurrence en perd ses plumes.     |
| Alliteration, assonance, and consonance | 107       | Weasleys’ Wildfire Whizbangs ( <i>Harry Potter</i> )   | Feuxfous Fuseboum   |
| Malapropisms                            | 35        | I’m Redd White, CEO of Bluecorp. You know, Corporate Expansion Official? ( <i>Phoenix Wright</i> ) | Je suis Redd White, le PDG de Bluecorp. Vous savez, Présence, Distinction et Grâce! |
| Others                                  | 25        | Snipperwhapper! ( <i>Phoenix Wright</i> )  | Vetit aporton!  |

absurd humour, humour related to the visual or historical context, humour related to a cultural or historical reference). We classified these instances according to their properties and annotated them as such, as the translation of absurd humour and cultural references is to be studied with different tools. These entries are not included in the tables.

We observed that there was significant overlap among the classes. The most problematic category was neologism, as almost any transformation of a common word was considered by our annotators to be a neologism. To address this issue, we decided to introduce our own classification and wordplay interpretation annotation scheme, which is described in §4.

## 4. Final annotation guidelines

For our Task 1, we annotated both phrase-based (puns) and term-based instances of wordplay in English and French (see §5). Following the SemEval-2017 pun task [25], we annotated each instance of wordplay according to its LOCATION and INTERPRETATION. For LOCATION, we mean the precise word(s) in the instance forming the wordplay, such as the ambiguous words of a punning joke.<sup>1</sup> INTERPRETATION means the explanation of the wordplay, which we do, for example, by providing the secondary meaning of a pun. To facilitate preprocessing, we do not use WordNet as in SemEval-2017 but rather introduce the notation described in Table 3.

We further annotated the data according to the following typologies:

<sup>1</sup>Unlike in the SemEval-2017 task, we simply list the word(s) in question rather than indicating their position within the instance.

**Table 2**

Statistics and examples for the initial classification of wordplay structure.

| Structure                               | # entries | English   | French   |
|---|-----------|---|--|
| Lexical structure (polysemy and idioms) | 337       | I used to be a train driver but I got sidetracked. (punning tweet)  | Avant, j'étais conducteur de train, mais j'ai changé de voie.  |
| Paronymy                                | 314       | I guess Lotta's in "lotta" trouble... ( <i>Phoenix Wright</i> )   | À mon avis, Eva, "eva" avoir des ennuis...   |
| Homophony                               | 161       | Weasleys' Wildfire Whizbangs ( <i>Harry Potter</i> )  | Feuxfous Fuseboum  |
| Homonymy                                | 54        | There's a large mustard-mine near here. And the moral of that is—The more there is of mine, the less there is of yours. ( <i>Alice's Adventures in Wonderland</i> ) | Il y a une bonne mine de moutarde près d'ici; la morale en est qu'il faut faire bonne mine à tout le monde!        |
| Syntactic or morphological structure    | 18        | I can't remember how to write 1, 1000, 51, 6 and 500 in Roman Numerals. IM LIVID. (punning tweet)   | J'ai oublié comment écrire 100, 1, 6 et 50 en chiffres romains, et même si cela m'agace fortement, je reste CIVIL. |

- HORIZONTAL/VERTICAL concerns the co-presence of source and target of the wordplay. In horizontal wordplay, both the source and the target of the wordplay are given:

**Example 4.1.** They're called lessons because they lessen from day to day.

In vertical wordplay, source and target are collapsed in a single occurrence:

**Example 4.2.** How do you make a cat drink? Easy: put it in a liquidizer.

- MANIPULATION TYPE:
  - Identity: source and target are formally identical, as in Example 4.2.
  - Similarity: as in Example 4.1: source and target are not perfectly identical, but the resemblance is obvious.
  - Permutation: the textual material is given a new order, as in anagrams or spoonerisms:
 

**Example 4.3.** Dormitory = dirty room
  - Abbreviation: an ad-hoc category for textual material where the initials form another meaning, as in acrostics or "funny" acronyms:

**Table 3**

Wordplay interpretation notation

|                            |   |
|----------------------------|---|
| $a / b$                    | Distinguishes the location from the interpretation and the different meanings of a wordplay:<br><i>meaning 1 (location) / meaning 2 (second meaning)</i>              |
| $a   b$                    | Separates the wordplay instances and their respective interpretations. An expression can contain several wordplay instances:<br><i>location 1   location 2</i>        |
| $a ( b )$                  | Specifies definitions or synonyms for each interpretation when location and interpretation are homographs:<br><i>meaning (synonym, hyperonym or brief definition)</i> |
| $a [ b ]$                  | Specifies comments like foreign language, anagram, palindrome etc.:<br><i>interpretation [anagram]</i>  |
| $a \{ b \}$                | Specifies the frame that activates the ambiguous word when a synonym or a short definition is not available:<br><i>meaning {frame activated by meaning}</i>           |
| $\langle a; \dots \rangle$ | Groups words from the same lexical field:<br><i>\langle word 1; word 2; word 3 \rangle</i>  |
| " a "                      | Indicates presence of an idiom:<br><i>"idiom"</i>   |
| $a \sim b$                 | Indicates several possible interpretations for an ambiguous word:<br><i>meaning 1 (interpretation 1) ~ meaning 2 (interpretation 2)</i>                               |
| $a + b$                    | Indicates that several words or syllables have been combined:<br><i>meaning 1 / meaning 1a + meaning 1b</i>   |
| $A / b$                    | Defines acronyms:<br><i>OWL /Ordinary Wizarding Level</i>   |
| $a \& b$                   | Shows when the wordplay relies on opposition:<br><i>location 1 &amp; location 2</i>   |

**Example 4.4.** BRAINS: Biobehavioral Research Awards for Innovative New Scientists

- Opposition: covers wordplay such as the antonyms *hot & ice / warms & freezing* in the following:

**Example 4.5.** Hot ice cream warms you up no end in freezing weather.

- MANIPULATION LEVEL: Most wordplay involves some kind of phonological manipulation, making SOUND our default category. Examples 4.1 and 4.2 involve a clear sound similarity or identity, respectively. Only if this category cannot be applied to the wordplay is the instance tagged with another level of manipulation. The next level to be considered is WRITING (as in Examples 4.3 and 4.4). If neither SOUND nor WRITING are manipulated, the level of manipulation is specified as OTHER. This level of manipulation may arise, for instance,

in chiasmuses:

**Example 4.6.** We shape our buildings, and afterwards our buildings shape us.

- **CULTURAL REFERENCE:** This is a binary (true/false) category. In order to understand some instances of wordplay, one has to be aware of some extra-linguistic factors.
- **CONVENTIONAL FORM:** Another binary category, this time indicating whether the wordplay occurs in a fixed form, such as a Tom Swifty (i.e., wellerism).
- **OFFENSIVE:** Another binary category, this time indicating whether the wordplay could be considered offensive. (This category was not evaluated in the pilot tasks.)

## 5. Corpus details

We constructed a parallel corpus of wordplay in English and French. Our data is twofold, containing phrase-based wordplay (puns) and term-based wordplay (mainly named entities).

### 5.1. Parallel corpus of puns

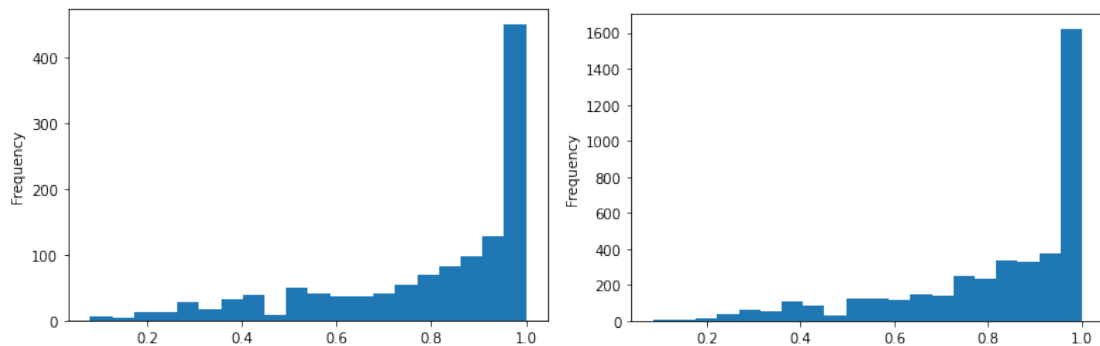
Our English corpus of puns is mainly based on that of the SemEval-2017 shared task on pun identification [25]. The original annotated dataset contains 3387 standalone English-language punning jokes, between 2 and 69 words in length, sourced from offline and online joke collections. Roughly half of the puns in the collection are “weakly” homographic (meaning that the lexical units corresponding to the two senses of the pun, disregarding inflections and particles, are spelled identically) while the other half are heterographic (that is, with lemmas spelled differently). The original annotation scheme is rather simple, indicating only the pun’s location within the joke, whether it is homographic or heterographic, and the two meanings of the pun (with reference to senses in WordNet [26]).

In order to translate this subcorpus from English into French, we applied a gamification strategy. More precisely, we organised a translation contest<sup>2</sup>. The contest was open to students but we also received multiple translations out of official ranking from professional translators and academics in translation studies. The results were submitted via Google Form. 47 participants submitted 3,950 translation of 500 puns coming from the SemEval-2017 dataset. We took first 250 puns in English from each of homographic and heterographic subsets. In the Google Form the homographic and heterographic puns were alternated. Each page of the Google Form contained 100 puns.

Besides this SemEval-derived data, we sourced further translation pairs from published literature and from puns translated by Master’s students in translation.

---

<sup>2</sup><https://www.joker-project.com/pun-translation-contest/>



**Figure 2:** Wordplay location normalised by text length for English (left); French (right)

We annotated our dataset according to the classification introduced in §9. The final annotated training set contains a total of 1772 distinct instances in English with 4753 corresponding French translations.

## 5.2. Parallel corpus of term-based wordplay

For this part of the corpus, we collected 1409 single terms in English containing wordplay from video games, advertising slogans, literature, and other sources [15] along with 1420 translations into French. Almost all translations are official ones but we have eleven additional ones proposed by our interns, Master’s students in translation.

Statistics on the annotated data are given in Tables 4 and 5. We furthermore noticed that the LOCATION is usually the last word in wordplay, as evidenced in Figure 2.

## 5.3. Training data

Our training data consists of 2078 wordplay instances in English and 2550 in French in the form of a list of translated wordplay instances. This data was provided as a JSON or CSV file with one fields for the unique ID of the instance, one for the text of the instance, and one each for the LOCATION, INTERPRETATION, HORIZONTAL/VERTICAL, MANIPULATION\_TYPE, MANIPULATION\_LEVEL, and CULTURAL\_REFERENCE annotations. Figure 3 shows an excerpt from the JSON file.

## 5.4. Test data

Our test data contains 3255 instances of wordplay in English from the SemEval-2017 pun task [25] and 4291 instances in French that we did not use for the training set. The test data was provided as a JSON or CSV file with only two fields – one of them a unique ID and the other the text of the instance. Figure 4 shows an excerpt of the JSON test data.

**Table 4**

Annotation statistics of puns.

| English  | French  |
|--|---|
| <ul style="list-style-type: none"> <li>• 1772 annotated instances               <ul style="list-style-type: none"> <li>- Vertical 1382</li> <li>- Horizontal 212</li> </ul> </li> <li>• MANIPULATION TYPE               <ul style="list-style-type: none"> <li>- Identity 894</li> <li>- Similarity 639</li> <li>- Opposition 42</li> <li>- Abbreviation 12</li> <li>- Permutation 7</li> </ul> </li> <li>• MANIPULATION LEVEL               <ul style="list-style-type: none"> <li>- Sound 1551</li> <li>- Writing 46</li> <li>- Other 2</li> </ul> </li> <li>• CULTURAL REFERENCE               <ul style="list-style-type: none"> <li>- False 1689</li> <li>- True 82</li> </ul> </li> <li>• CONVENTIONAL FORM               <ul style="list-style-type: none"> <li>- False 1604</li> <li>- True 167</li> </ul> </li> <li>• OFFENSIVE               <ul style="list-style-type: none"> <li>- Sexist 9</li> <li>- Possibly 7</li> <li>- Racist 2</li> <li>- Other 1</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>• 4753 annotated instances               <ul style="list-style-type: none"> <li>- Vertical 4400</li> <li>- Horizontal 320</li> </ul> </li> <li>• MANIPULATION TYPE               <ul style="list-style-type: none"> <li>- Identity 2970</li> <li>- Similarity 1672</li> <li>- Opposition 51</li> <li>- Permutation 17</li> <li>- Abbreviation 9</li> </ul> </li> <li>• MANIPULATION LEVEL               <ul style="list-style-type: none"> <li>- Sound 4540</li> <li>- Writing 179</li> <li>- Other 4</li> </ul> </li> <li>• CULTURAL REFERENCE               <ul style="list-style-type: none"> <li>- False 4665</li> <li>- True 88</li> </ul> </li> <li>• CONVENTIONAL FORM               <ul style="list-style-type: none"> <li>- False 4665</li> <li>- True 88</li> </ul> </li> <li>• OFFENSIVE               <ul style="list-style-type: none"> <li>- Sexist 21</li> <li>- Possibly 6</li> <li>- Racist 4</li> <li>- Other 1</li> </ul> </li> </ul> |

The prescribed output format is similar to the training data format, but with the addition of the fields RUN\_ID (to uniquely identify the participating team, pilot task, and run number), MANUAL (to indicate whether the output annotations are produced by a human or a machine), and OFFENSIVE (per our annotation scheme).

## 6. Preliminary results on wordplay perception

We carried out a preliminary analysis of wordplay perception based on the French data issued from the translation contest. A student in linguistics, a French native

**Table 5**

Annotation statistics of wordplay in named entities.

| English  | French   |
|--|--|
| <ul style="list-style-type: none"> <li>• 1409 annotated instances               <ul style="list-style-type: none"> <li>– Vertical 1408</li> <li>– Horizontal 1</li> </ul> </li> <li>• MANIPULATION TYPE               <ul style="list-style-type: none"> <li>– Similarity 606</li> <li>– Identity 441</li> <li>– Abbreviation 340</li> <li>– Permutation 17</li> <li>– Opposition 1</li> </ul> </li> <li>• MANIPULATION LEVEL               <ul style="list-style-type: none"> <li>– Sound 1402</li> <li>– Writing 7</li> </ul> </li> <li>• CULTURAL REFERENCE               <ul style="list-style-type: none"> <li>– False 1361</li> <li>– True 48</li> </ul> </li> <li>• CONVENTIONAL FORM               <ul style="list-style-type: none"> <li>– NOT APPLICABLE</li> </ul> </li> <li>• OFFENSIVE               <ul style="list-style-type: none"> <li>– NOT IDENTIFIED</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>• 1420 annotated instances               <ul style="list-style-type: none"> <li>– Vertical - 1419</li> <li>– Horizontal - 1</li> </ul> </li> <li>• MANIPULATION TYPE               <ul style="list-style-type: none"> <li>– Similarity 775</li> <li>– Identity 415</li> <li>– Abbreviation 211</li> <li>– Permutation 15</li> <li>– Opposition 1</li> </ul> </li> <li>• MANIPULATION LEVEL               <ul style="list-style-type: none"> <li>– Sound 1411</li> <li>– Writing 9</li> </ul> </li> <li>• CULTURAL REFERENCE               <ul style="list-style-type: none"> <li>– False 1344</li> <li>– True 76</li> </ul> </li> <li>• CONVENTIONAL FORM               <ul style="list-style-type: none"> <li>– NOT APPLICABLE</li> </ul> </li> <li>• OFFENSIVE               <ul style="list-style-type: none"> <li>– NOT IDENTIFIED</li> </ul> </li> </ul> |

speaker, applied a score between 0 and 5 on a Likert scale [27] to evaluate joke humorousness. For the 149 annotated wordplay instances in French, the average humorousness score was 4.6. Among the annotated examples, there were several wellerisms:

**Question-answer** (25 in total). This type of wellerism refers to bipartite jokes with the form of a short dialogue: a question followed by an answer.

**Example 6.1.** Qu'est-ce que tu fais sur une île déserte ? Tu trouves une cuillère et tu l'attaques.

**Old soldiers never die** (7 in total) These wellerisms are variations on a catchphrase, with the original version being "Old soldiers never die, they simply fade away."

```
[
  {
    "ID": "noun_1063",
    "WORDPLAY": "Elimentaler",
    "LOCATION": "Elimentaler",
    "INTERPRETATION": "Emmental (cheese) + Eliminator",
    "HORIZONTAL/VERTICAL": "vertical",
    "MANIPULATION_TYPE": "Similarity",
    "MANIPULATION_LEVEL": "Sound",
    "CULTURAL_REFERENCE": false,
    "CONVENTIONAL_FORM": false,
    "OFFENSIVE": null
  },
  {
    "ID": "pun_341",
    "WORDPLAY": "Geologists can be sedimental about their work.",
    "LOCATION": "sedimental",
    "INTERPRETATION": "sentimental/sediment",
    "HORIZONTAL/VERTICAL": "vertical",
    "MANIPULATION_TYPE": "Similarity",
    "MANIPULATION_LEVEL": "Sound",
    "CULTURAL_REFERENCE": false,
    "CONVENTIONAL_FORM": false,
    "OFFENSIVE": null
  }
]
```

**Figure 3:** Excerpt of training data (JSON format)

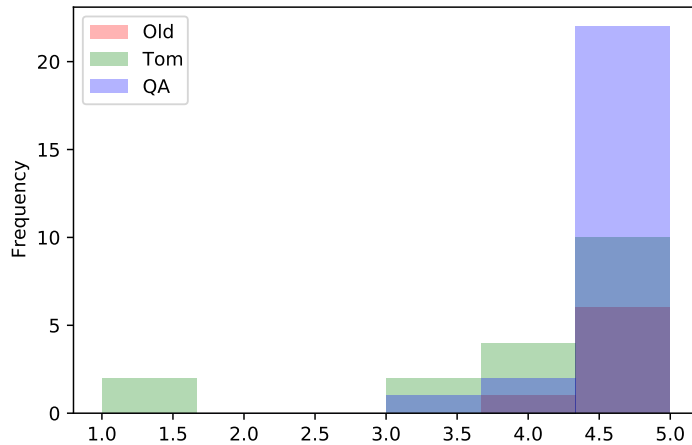
```
[
  {
    "ID": "noun_1",
    "WORDPLAY": "Ambipom"
  },
  {
    "ID": "het_1011",
    "WORDPLAY": "These are my parents, said Einstein relatively"
  }
]
```

**Figure 4:** Excerpt of test data (JSON format)

**Example 6.2.** Les vieux adeptes de saut à l'élastique ne meurent jamais : ils savent toujours rebondir.

**Tom Swifty** (18 in total) These are wellerisms with a phrase in which a quoted sentence is linked by a pun to the manner in which it is attributed. The





**Figure 5:** Histogram of wellerism hilariousness

standard form is for the quoted sentence to be first, followed by the description of the act of speaking of the conventional speaker, Tom.

**Example 6.3.** “Pourquoi est-ce que je ne me vois pas dans ce miroir?” fit Tom sans réfléchir.

Figure 5 presents the histogram of wellerism humorousness, with free-text comments for 73 jokes reproduced in Table 6. As is clear from the figure, Tom Swifties were generally not considered funny, while the highest scores were given to question-answer wellerisms. These results are somehow opposite to the generated joke humorousness in [28]. Although this opposition seems obvious due to the method used for wellerism generation in [28], further analysis is needed as the annotators were different and humour perception depends on multiple social factors, including genre and age.

Looking at the manually constructed data, we noticed that in a few instances, style shift in the translation of the pun could pose an issue. Consider the following pair:

**Example 6.4.** I phoned the zoo but the lion was busy.  
J’ai appelé le zoo mais on m’a dit phoque you.

The French translation includes a vulgarism, with a pun across languages (fuck/phoque). This was considered a very successful translation, but would clearly be an inappropriate translation in many contexts. A number of other examples that we could spot introduced strong stereotyping that could be construed as offensive, in contrast to the original.

We decided to annotate the data for those style shifts that introduced in the translation a form of humour relying on vulgarism or stereotyping. In doing so,

**Table 6**

Free comment statistics

| Free comment       | # wordplay |
|--------------------|------------|
| ?                  | 2          |
| fun                | 3          |
| boosting           | 1          |
| funny              | 25         |
| hard               | 2          |
| dynamic            | 3          |
| intellectual       | 11         |
| literary           | 1          |
| cute               | 12         |
| not funny          | 3          |
| no more            | 1          |
| sexist             | 1          |
| sexist and sexual? | 1          |
| sexual             | 3          |
| serious            | 4          |

another issue became evident: an additional bias may be introduced in the data due to the French language. Consider the following pair:

**Example 6.5.** Old Quilters never die, they just go under cover.

Les vieilles tricoteuses ne meurent jamais, elles recousent les morceaux.

French is more strongly gendered than English. As many French speakers still consider the use of masculine a default, this translation introduces a stereotype by using a feminine translation for the word *knitter* (*tricoteuse*). However, using the masculine form only, as a *default gender*, also raises questions in a context where the current evolution of the language seems to go against that usage [29].

## 7. Classifications proposed by participants

The JOKER participants suggested new classifications of wordplay in an attempt to overcome issues with the existing classifications.

Delarche [30] distinguishes polysemic constructs and letter-based constructs – e.g., wordplay based on selections, permutations, repetitions or suppression of letters such as acronyms, acrostics, lipograms, palindromes, pangrams. We should admit that this distinction, although not tested, seems promising for tasks of wordplay generation and translation. He describes in details acronyms and acrostics. These types of wordplay are missing from our corpus.

Delarche [30] also differentiates single pivotal keyword polysemic constructs from repeated keyword with different meanings, which is basically similar to the categories *HORIZONTAL/VERTICAL* that we ourselves defined [15].

A. Digue and P. Campen tried to make the JOKER classification more precise by introducing a clear distinction between *Sound/Writing/Both* for *VERTICAL* wordplay and *Sound/Writing/Both/Other* for *HORIZONTAL* one. They also demonstrated the non-existence of certain combinations of the JOKER categories.

## 8. Methods used by the participants

Five teams participated in Pilot Task 1: FAST\_MT [31], eBIHAR [32], Cecilia [33], Agnieszka, and Hakima [34].

The Cecilia and Agnieszka teams applied the Google T5 model [35] via the SimpleT5 library<sup>3</sup>. The Google T5 (Text-To-Text Transfer Transformer) model is based on the transfer learning with a unified text-to-text transformer [35]. Agnieszka submitted a run without a paper, though the team notified the JOKER organisers of the method they used.

eBIHAR applied a polynomial naive Bayesian classifier and logistic regression to classify and predict text (with and without preprocessing) on the bag-of-words and TF-IDF representations.

Hakima applied Jurassic-1, a model of the first generation in a series of large language models trained and made widely accessible by AI21 Labs<sup>4</sup> [36]. Jurassic-1 is an auto-regressive language model based on the decoder module of the Transformer architecture [37] with the modifications similar to GPT-3 proposed by Radford, Wu, Child, Luan, Amodei, and Sutskever [38].

One team submitted a run after the official deadline and so its results are not presented in our workshop overview paper [15].

Delarche [30] suggested interesting heuristic-based deterministic algorithmic filters which might be useful for specific types of wordplay, though these algorithms were not implemented and therefore not tested.

## 9. Evaluation metrics

We preprocessed runs to lowercase and trim the values. For the English subcorpus, the labels for LOCATION and INTERPRETATION were provided for puns from the original dataset [25]. All wordplay instances from this dataset were considered to be VERTICAL with manipulation type SOUND. HOMOGRAPHIC puns were attributed the IDENTITY label while HETEROGRAPHIC puns were classified as SIMILARITY manipulation type. We report the absolute values of true labels submitted by the participants.

We discarded all INTERPRETATION values that were equal to the LOCATION fields as we considered this to be insufficient.

---

<sup>3</sup><https://github.com/Shivanandroy/simpleT5>

<sup>4</sup><https://studio.ai21.com/>

**Table 7**

Scores of participants' runs for Pilot 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<br>_wp_preprocessed |          | 536         | 2437         |

In recognition of the fact that there may be slightly different but equally valid INTERPRETATION annotations, for evaluation we retained only the high-level annotation (by removing everything in brackets, parentheses, etc.). We downcased, tokenised, and lemmatised this high-level annotation with the aid of regular expressions and the NLTK WordNetLemmatizer.<sup>5</sup> We then compared the set of lemmas generated by participants with our own annotations.

## 10. Results

All together, four teams submitted eight runs for the English dataset. The eBIHAR team also submitted one run in French. The release of the French dataset was delayed and we also updated the English dataset during the competition. The FAST\_MT team submitted runs both for the first release of English dataset and the updated one. The Agnieszka team submitted only partial runs for LOCATION. The results for the participants are given in Table 7.

All participants, except the Agnieszka team which did not submit predictions for MANIPULATION LEVEL, successfully predicted all classes. However, this success might be explained by the nature of our data, as in the test set the only class was SOUND.

The teams Cecilia, FAST\_MT, and Agnieszka demonstrated fairly good results for LOCATION. However, as previously noted, in our dataset the majority of instances had the wordplay located at the last word.

Only the FAST\_MT team succeeded in INTERPRETATION prediction for the first data release. For this first run, our annotation coincides with that of the submission in 597 cases; it differs for 61. These differences are, in the majority of cases, not errors but differences in the presentation or human interpretation. The first dataset contained a lot of named entities from popular anime, movies, and video games (e.g., Pokemon), unlike the updated data set. FAST\_MT had gathered raw data from various websites explaining puns in Pokemon names and trained their model on it. We should

<sup>5</sup>[https://www.nltk.org/\\_modules/nltk/stem/wordnet.html](https://www.nltk.org/_modules/nltk/stem/wordnet.html)

acknowledge that some annotations provided by FAST\_MT were more detailed than ours. For the updated dataset, FAST\_MT's predictions for LOCATION are identical to those of INTERPRETATION. Only one run, Cecilia's run 5, was successful for this dataset with 441 correct results.

We do not provide results for other binary classes; since our data was unbalanced with regard to these categories, the submitted results always provided negative labels.

## 11. Conclusion

We introduced the JOKER track at CLEF 2022, consisting of a workshop and associated pilot tasks on automatic wordplay analysis and translation. Our primary goal is to build parallel data and evaluation metrics for detecting, locating, interpreting, and translating wordplay in order to wوزه a step forward to the automation of wordplay analysis.

We surveyed existing classifications and we present the data we initially annotated according to two well-known classifications from the literature. However, we were obliged to abandon these classifications as the first one contains overlapping classes while another is not expressive enough. Therefore, we introduced a new classification of wordplay which aims to handle the issues of the classifications from the literature.

We manually classified wordplay in English and French according to our categories. Our data was used to organise Pilot Task 1: Classify and Explain Instances of Wordplay. Four teams submitted official runs for the Pilot Task 1; one team submitted a run after the deadline.

Participants succeeded in wordplay location, but the interpretation tasks raised difficulties. The binary classes HORIZONTAL/VERTICAL, CONVENTIONAL\_FORM, CULTURAL\_REFERENCE, OFFENSIVE, MANIPULATION\_LEVEL were unbalanced, provoking very high but meaningless scores. However, these binary classifications were not the focus of our research. We plan to perform a more detailed study of wordplay perception, including humorousness, and offensiveness, as well as a free category study.

It should be kept in mind that our data consists mainly of puns and portmanteaux, which may make our classification too not expressive enough. Participants proposed new wordplay classifications or tried to improve upon ours. In the future, we will use this feedback to revise our classification in order to improve its expressiveness.

Further details on the other pilot tasks and the submitted runs can be found in the CLEF CEUR proceedings [39]. The overview of the entire JOKER track can be found in the LNCS proceedings [15]. Additional information on the track is available on the JOKER website: <http://www.joker-project.com/>

## 12. Authors' contribution

The general framework was proposed by L. Ermakova. The initial annotation guide was proposed by L. Ermakova and G. Le Corre. The new classification was introduced by F. Regattin and adjusted by L. Ermakova, T. Miller, J. Boccou, A. Digue, A. Damoy with participation of C. Borg and S. Araújo. O. Puchalski worked on the initial data annotation. The interpretation annotation is an extension of the work of T. Miller and was proposed by L. Ermakova and adjusted J. Boccou, A. Digue, A. Damoy, and Paul Campen. J. Boccou, A. Digue, and A. Damoy annotated data. A.-G. Bosser worked on the perception aspects and general organisation of the evaluation campaign. Evaluation results were obtained and described by L. Ermakova. J. Boccou wrote the first draft of the annotation guidelines. S. Araújo, G. Le Corre and F. Regattin wrote the state of the art.

## Acknowledgments

This work has been funded in part by the National Research Agency under the program *Investissements d'avenir* (Reference ANR-19-GURE-0001) and by the Austrian Science Fund under project M2625-N31. JOKER is supported by *La Maison des sciences de l'homme en Bretagne*.

We thank Adrien Couaillet and Ludivine Grégoire for their participation in data collection, annotation and adjustment of classification guidelines. We also thank Elise Mathurin for co-supervising interns in translation as well as Alain Kerhervé who supported the project.

We would like also thank other JOKER organisers: Anne-Gwenn Bosser, Claudine Borg, Fabio Regattin, Gaelle Le Corre, Elise Mathurin, Silvia Araujo, Monika Bokinieć, Górg Mallia, Gordan Matas, Mohamed Saki, Benoît Jeanjean, Radia Hannachi, Danica Škara, and other PC members: Grigori Sidorov, Victor Manuel Palma Preciado, and Fabrice Antoine.

We thank Eric Sanjuan who provided resources for data management.

## References

- [1] L. Laineste, P. Voolaid, Laughing across borders: Intertextuality of internet memes, *The European Journal of Humour Research* 4 (2017) 26–49.
- [2] R. A. Martin, *The Psychology of Humor: An Integrative Approach*, Academic Press, 2006. doi:10.5860/choice.45-2902.
- [3] T. Jiang, H. Li, Y. Hou, Cultural differences in humor perception, usage, and implications, *Frontiers in Psychology* 10 (2019) 141–156. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6361813/>. doi:10.3389/fpsyg.2019.00123.
- [4] S. Attardo (Ed.), *The Linguistics of Humor: An Introduction*, Oxford Scholarship Online, 2020. doi:10.1093/oso/9780198791270.001.0001.
- [5] D. Chiaro, *Humor and Translation*, Routledge, 2017, p. 16.

- [6] M. J. Veiga, Linguistic mechanisms of humour subtitling, in: IV Forum for Linguistic Sharing, 2009, pp. 1–14. URL: <https://clunl.fcsh.unl.pt/wp-content/uploads/sites/12/2017/07/linguistic-mechanisms-of-humour-subtitling.pdf>.
- [7] L. Ermakova, T. Miller, O. Puchalski, F. Regattin, É. Mathurin, S. Araújo, A.-G. Bossler, C. Borg, M. Bokiniec, G. L. Corre, B. Jeanjean, R. Hannachi, Ğ. Mallia, G. Matas, M. Saki, CLEF Workshop JOKER: Automatic Wordplay and Humour Translation, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), *Advances in Information Retrieval*, volume 13186 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2022, pp. 355–363. doi:10.1007/978-3-030-99739-7\_45.
- [8] A. Zirker, E. Winter-Froemel (Eds.), *Wordplay and Metalinguistic/Metadiscursive Reflection: Authors, Contexts, Techniques, and Meta-Reflection*, volume 2015 of *English and American Studies in German*, 2015.
- [9] Z. Yu, J. Tan, X. Wan, A neural approach to pun generation, in: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, volume 1, Association for Computational Linguistics, 2018, pp. 1650–1660. URL: <https://aclanthology.org/P18-1153>. doi:10.18653/v1/P18-1153.
- [10] A. Jaiswal, Monika, A. Mathur, Prachi, S. Mattu, Automatic humour detection in tweets using soft computing paradigms, 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (2019) 172–176.
- [11] T. Miller, C. F. Hempelmann, I. Gurevych, SemEval-2017 Task 7: Detection and interpretation of English puns, in: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017, pp. 58–68. doi:10.18653/v1/S17-2005.
- [12] T. Miller, M. Turkovic, Towards the automatic detection and identification of english puns, *The European Journal of Humour Research* 4 (2016) 59–75. doi:10.7592/EJHR2016.4.1.miller.
- [13] A. Mittal, Y. Tian, N. Peng, Ambipun: Generating humorous puns with ambiguous context, ArXiv abs/2205.01825 (2022).
- [14] R. Sharma, S. Shekhar, An automatic pun word identification framework for code mixed text, in: *Proceedings of the 5th International Conference on Information Systems and Computer Networks (ISCON)*, 2021, pp. 1–5.
- [15] L. Ermakova, T. Miller, F. Regattin, A.-G. Bossler, E. Mathurin, G. L. Corre, S. Araújo, J. Boccou, A. Digue, A. Damoy, B. Jeanjean, Overview of JOKER@CLEF 2022: Automatic Wordplay and Humour Translation workshop, in: A. Barrón-Cedeño, G. Da San Martino, M. Degli Esposti, F. Sebastiani, C. Macdonald, G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, N. Ferro (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022)*, volume 13390 of *LNCS*, 2022.
- [16] D. Delabastita, Introduction to the special issue on wordplay and translation, *The Translator: Studies in Intercultural Communication* 2 (1996) 1–22. doi:10.1080/13556509.1996.10798970.

- [17] H. Gottlieb, Anglicisms and translation, in: G. Anderman, M. Rogers (Eds.), *In and Out of English: For Better, For Worse*, Multilingual Matters, 2005, pp. 161–184. doi:10.21832/9781853597893-014.
- [18] D. Chiaro, *The Language of Jokes. Analyzing Verbal Play*, Routledge, London, 1992.
- [19] M. Giorgadze, Linguistic features of pun, its typology and classification, *European Scientific Journal* 10 (2014). URL: <https://eujournal.org/index.php/esj/article/view/4819>.
- [20] W. Redfern, *Puns*, Blackwell, 1985. doi:10.1177/007542428702000114.
- [21] H. Gottlieb, "You got the picture?". On the Polysemiotics of subtitling wordplay, St. Jerome Publishing, 1997, pp. 207–232.
- [22] M. Mastonen, *Translating Wordplay: a Case Study on the Translation of Wordplay in Terry Pratchett's Soul Music*, Master's thesis, School of Languages and Translation Studies, Faculty of Humanities, University of Turku, 2016. URL: <https://www.utupub.fi/bitstream/handle/10024/146151/MustonenMarjo.pdf>.
- [23] D. Delabastita, *There's a Double Tongue: an Investigation into the Translation of Shakespeare's Wordplay, with Special Reference to Hamlet*, Rodopi, Amsterdam, 1993.
- [24] Y. Chuandao, English pun and its classification, *Language in India* 5 (2005). URL: <http://www.languageinindia.com/april2005/englishpun1.html>.
- [25] T. Miller, C. F. Hempelmann, I. Gurevych, SemEval-2017 Task 7: Detection and interpretation of English puns, in: *Proceedings of the 11th International Workshop on Semantic Evaluation*, 2017, pp. 58–68. doi:10.18653/v1/S17-2005.
- [26] C. Fellbaum (Ed.), *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, MA, 1998.
- [27] R. Likert, A technique for the measurement of attitudes, *Archives of Psychology* 22 140 (1932) 55–55.
- [28] L. Glémarec, A.-G. Bossier, L. Ermakova, Generating Humorous Puns in French, in: *Proceedings of the Working Notes of CLEF 2022 – Conference and Labs of the Evaluation Forum*, Bologna, Italy, September 5th to 8th, 2022, CEUR Workshop Proceedings, CEUR-WS.org, Bologna, Italy, 2022, p. 8.
- [29] E. Viennot, *Le langage inclusif: pourquoi, comment.*, Les Éditions iXe, 2020.
- [30] M. Delarche, A translation-oriented categorisation of wordplays, in: *Proceedings of the Working Notes of CLEF 2022 – Conference and Labs of the Evaluation Forum*, Bologna, Italy, September 5th to 8th, 2022, CEUR Workshop Proceedings, CEUR-WS.org, Bologna, Italy, 2022, p. 6.
- [31] F. Dhanani, M. Rafi, M. A. Tahir, FAST\_MT participation for the JOKER CLEF-2022 automatic pun and human translation tasks, in: *Proceedings of the Working Notes of CLEF 2022 – Conference and Labs of the Evaluation Forum*, Bologna, Italy, September 5th to 8th, 2022, CEUR Workshop Proceedings, CEUR-WS.org, Bologna, Italy, 2022, p. 14.
- [32] A. Epimakhova, Using machine learning to classify and interpret wordplay, in: *Proceedings of the Working Notes of CLEF 2022 – Conference and Labs of the Evaluation Forum*, Bologna, Italy, September 5th to 8th, 2022, CEUR Workshop



- Proceedings, CEUR-WS.org, Bologna, Italy, 2022, p. 6.
- [33] L. Glemarec, Use of SimpleT5 for the CLEF workshop JokeR: Automatic Pun and Humor Translation, in: Proceedings of the Working Notes of CLEF 2022 – Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th to 8th, 2022, CEUR Workshop Proceedings, CEUR-WS.org, Bologna, Italy, 2022, p. 11.
- [34] H. Arroubat, CLEF Workshop: Automatic Pun and Humour Translation Task, in: Proceedings of the Working Notes of CLEF 2022 – Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th to 8th, 2022, CEUR Workshop Proceedings, CEUR-WS.org, Bologna, Italy, 2022.
- [35] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, *Journal of Machine Learning Research* 21 (2020) 1–67. URL: <http://jmlr.org/papers/v21/20-074.html>.
- [36] O. Lieber, O. Sharir, B. Lentz, Y. Shoham, Jurassic-1: Technical Details and Evaluation, White paper, AI21 Labs, 2021. URL: [https://uploads-ssl.webflow.com/60fd4503684b466578c0d307/61138924626a6981ee09caf6\\_jurassic\\_tech\\_paper.pdf](https://uploads-ssl.webflow.com/60fd4503684b466578c0d307/61138924626a6981ee09caf6_jurassic_tech_paper.pdf).
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *arXiv:1706.03762 [cs]* (2017). URL: <http://arxiv.org/abs/1706.03762>.
- [38] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, Language Models Are Unsupervised Multitask Learners, Technical report, OpenAI, 2019. URL: [https://cdn.openai.com/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf).
- [39] G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022: Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, CEUR-WS.org, 2022.