Forbidden FRUIT is the Sweetest: An Annotated Tweet Corpus for French Unfrozen Idioms Identification

Julien Bezançon[1,2], **Félix Alié**[2], **Antoine Gautier**[1],
Marceau Hernandez[1,2], **Gaël Lejeune**[1,2]
[1]STIH, Sorbonne Université, Paris, France
[2]CERES, Sorbonne Université, Paris, France
firstname.lastname@sorbonne-universite.fr

Abstract

Multiword expressions (MWEs) are a key area of interest in NLP, studied across various languages and inspiring the creation of dedicated datasets and shared tasks such as PARSEME. Puns in multiword expressions (PMWEs) can be described as MWEs that have been "unfrozen" to acquire a new meaning or create wordplay. Unlike MWEs, they have received little attention in NLP, mainly due to the lack of resources available for their study. In this context, we introduce the French Unfrozen Idioms in Tweets (FRUIT) corpus, a dataset of tweets spanning three years and comprising 60,617 tweets containing both MWEs and PMWE candidates. We first describe the process of constructing this corpus, followed by an overview of the manual annotation task performed by three experts on 600 tweets, achieving an inter-annotator agreement score α up to 0.83. Insights from this manual annotation process were then used to develop a Game With A Purpose (GWAP) to annotate more tweets from the FRUIT corpus. This GWAP aims to enhance players' understanding of MWEs and PMWEs. Currently, 13 players made 2,206 annotations on 931 tweets, reaching an α score of 0.70. In total, 1,531 tweets from the FRUIT corpus have been annotated.

1 Introduction

Multiword Expressions (MWEs) have long posed a significant challenge in Natural Language Processing, sometimes referred to as a "pain in the neck" (Sag et al., 2002). The term MWE corresponds to a large span of linguistic objects, more or less subject to variations and with a certain degree of idiomaticity at the lexical, syntactic, semantic, pragmatic and/or statistical levels (Baldwin and Kim, 2010). Constant et al. (2017) describe them as both idiosyncratic and pervasive across different languages. MWEs are valuable not only for linguistic analysis but also for improving NLP tasks such as Machine Translation.

Wordplays and puns created from MWEs (hereafter PMWEs) can be described as MWEs that have undergone lexical, syntactic, semantic and/or pragmatic changes to create a wordplay. Their idiomatic status has been broken, leading to the emergence of a new meaning (Eline and Zhu, 2014). In linguistics, this phenomenon is often referred to as "*défigement*" (FR, "unfreezing"), which is often found in French linguistic literature. Mejri (2013) claims that the underlying MWE should always remain identifiable in a PMWE. Therefore, the MWE (1) is still recognisable in the PMWE (2).

1. *Tu quoque mi **fili*** (Latin, you too, my **son**)
2. *Tu quoque mi **chili*** (Latin, you too, my **chili**)

PMWE studies in NLP present several interests: (I) it can help to characterise MWEs by their productivity in wordplay (Lecler, 2006), (II) it allows the real-time detection of wordplays and even MWEs (Haßler and Hümmer, 2005; Cusimano, 2015) and (III) they shed light on the cognitive processes that allow human speakers to recognise these particular MWEs. We argue that such a study could also benefit MWEs recognition in NLP as PMWEs share the same linguistic challenges, such as idiomaticity across multiple levels, making them particularly challenging for NLP tasks like Machine Translation.

In this paper, we introduce the French Unfrozen Idioms in Tweets corpus (FRUIT), which consists of 60,617 tweets collected for the identification of French PMWEs. To our knowledge, no previous effort has been made to annotate PMWEs or create a dedicated corpus for them. The FRUIT corpus builds upon and expands an existing Twitter (now X) dataset (Bezançon and Lejeune, 2023). Section 3 details the corpus construction and the methodology for identifying PMWEs. We then introduce two annotation tasks:

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 70–85
July 5–10, 2020.

Manual Annotation Task Three experts in NLP and linguistics annotated 600 tweets containing potential MWEs and PMWEs, highlighting challenges in the identification of these entities, which we discuss in Section 6. The results of this annotation are available on GITHUB[1].

Annotation through a GWAP Using insights from the manual annotation task, we designed a GWAP to facilitate large-scale annotation of MWEs and PMWEs by a broader audience. The source code of this GWAP is available on GITHUB[2].

Through these annotation tasks, we aim to assess the difficulty of identifying both MWEs and PMWEs in tweets, combining expert knowledge with a gamified approach to enable non-expert contributors to participate in the annotation process. We provide the scripts used for tweet collection, along with all tweet IDs, in a dedicated GITHUB repository[3].

2 Related Work

MWE Identification As explained by Constant et al. (2017), MWE processing involves two main tasks: (i) discovery (ii) identification. Discovery involves detecting and adding MWEs to a lexicon, whereas identification focuses on automatically annotating MWEs in text. MWE identification is made very difficult by the evasive nature of MWEs (Geeraert et al., 2018). Savary et al. (2019) claims that without the creation of syntactic lexicons and at least some morphosyntactic information, we will not make significant progress on this task. Various approaches have been explored to build such lexicons, including crowdsourcing (Ramisch et al., 2016) and gamified platforms (Krstev and Savary, 2017; Fort et al., 2018, 2020). The PARSEME shared tasks (Savary et al., 2017) further demonstrate the community's commitment to improving MWE processing. As with MWEs, we believe that the creation of dedicated resources is a major challenge for identifying PMWEs.

GWAPs GWAPs (Games With A Purpose) correspond to games designed to let the machine learn from human inputs (Lafourcade et al., 2015). They have been widely used in NLP, particularly for resource creation (Lafourcade, 2007) and annotation (Hiebel et al., 2024; Madge et al., 2019). GWAPs offer several advantages: (i) they attract different types of players, such as the ones identified by Bartle (1996) and (ii) they provide an efficient alternative to traditional crowdsourcing methods (Fort et al., 2011; Fort, 2022). GWAPs have been successfully applied to MWE annotation, as demonstrated by RIGORMORTIS (Fort et al., 2020).

Wordplays While wordplay has been studied to some extent in NLP — particularly through shared tasks such as JOKER-CLEF (Ermakova et al., 2022, 2023, 2024) or the SEMEVAL tasks (Miller et al., 2017) — PMWEs remain largely unexplored. However, like Wordplays, PMWEs present unique challenges, both in terms of understanding linguistic creativity (Partington, 2009) and generating computationally creative text (Valitutti et al., 2013).

3 Building a French Tweets Corpus Containing PMWEs

3.1 Getting PMWEs Candidates

We compiled a list of 216 French MWEs to query the TWITTER API over a three-years period (from 2020 to 2023), yielding a dataset of 3,369,636 tweets. These MWEs were manually selected by four researchers specializing in NLP or linguistics. The only selection criterion was the conventionality of a MWE. Conventionalized MWEs tend to have a non-compositional meaning and are commonly recognized by speakers of a given language (Nunberg et al., 1994). Among these MWEs, we find (i) advertising slogans, (ii) famous quotes, (iii) movie catchphrases and (iv) other types of MWEs:

- (i) "*C'est le second effet Kisscool*" ("it's the second Kisscool effect", French advertising slogan for a chewing-gum brand)
- (ii) "*Travailler plus pour gagner plus*" ("work more to earn more", Nicolas Sarkozy, 2007)
- (iii) "*Dans l'espace, personne ne vous entend crier*" ("in space, no one will hear you scream", Alien movie catchphrase, 1979)
- (iv) "*Au bout du rouleau*" ("At the end of the rope")

Each tweet of this corpus is linked to the MWE that prompted its extraction (hereafter *seed*). Consequently, every tweet has some likelihood of con-

[1] https://github.com/JulienBez/ForbiddenFrUIT
[2] https://github.com/CERES-Sorbonne/Defricheur
[3] https://github.com/JulienBez/FrUIT

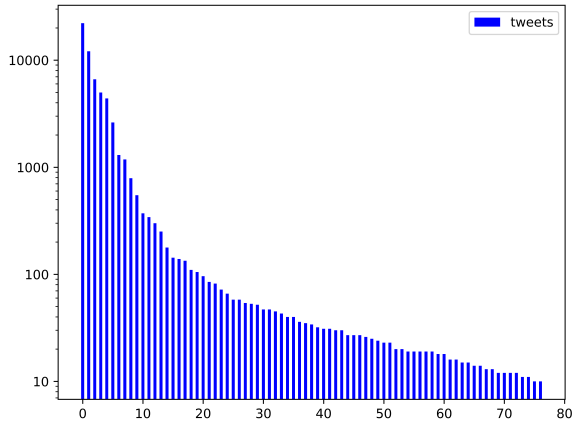

Figure 1: Logscale Zipf-like distribution of tweets per *seeds* in our corpus.

taining either a MWE or a PMWE, as it shares at least one word with its *seed*.

3.2 Filtering Steps

To retain only the most relevant tweets, we applied a three-step filtering process: (i) we discarded any tweet containing less than 50 % of the words of its corresponding *seed* (without preprocessing), (ii) we filtered out duplicates (tweets with identical IDs or texts) and (iii) we excluded tweets associated with *seeds* that appeared in fewer than ten tweets. This final step ensured that we retained only the most productive seeds.

After filtering, 60,617 tweets and 77 *seeds* remained. Figure 1 shows that the top ten *seeds* generated 86.51 % (56,769 tweets) of our dataset.

3.3 Asserting the Presence of PMWE Candidates

To complete the corpus creation, we aimed to verify the presence of PMWE candidates. To achieve this goal, we applied the algorithm introduced in (Bezançon and Lejeune, 2023). This algorithm uses token-level alignments between a MWE and a sentence to extract PMWE candidates, as illustrated in Table 1. It then ranks candidates for each MWE according to a cosine similarity score, measuring how closely a candidate resembles the original MWE. The higher the score, the closer a candidate is to a MWE (see Appendix A.1).

When comparing the MWE "*que la force soit avec toi*" ("May the force be with you", Stars Wars franchise) with the sequence "*que la force* ***ouvrière*** *soit avec toi*" ("May the **worker** force be with you"), found in a tweet, we observe the insertion of the word "*ouvrière*", creating the term "*Force Ouvrière*" ("worker force"), which is the name of a labor union in France. This change is captured in the alignment. Table 2 shows an example of the ranking obtained with this algorithm with the MWE "*que la force soit avec toi*". We used this algorithm to prioritize tweets most likely to contain PMWEs for annotation in Section 4.

que	*la*	*force*	*-*	*soit*	*avec*	*toi*
que	*la*	*force*	***ouvrière***	*soit*	*avec*	*toi*

Table 1: Token level alignment between the MWE "*que la force soit avec toi*" (may the force be with you) and a PMWE candidate.

Candidate							Score
que la	-	force	du ×2	soit	-	avec toi	0.85
que la	-	force	-	soit	tjrs	avec toi	0.83
que la	-	force	update	soit	-	avec toi	0.83
que la	-	force	rhétorique	soit	-	avec toi	0.83
que la	-	force	tranquille	soit	-	avec toi	0.83
que la	-	force	-	soit	toujours	avec toi	0.83
que la	-	force	marocaine	soit	-	avec toi	0.83
que la	vraie	force	-	soit	-	avec toi	0.83
que la	tri	force	-	soit	-	avec toi	0.81
que la	-	force	ouvrière	soit	-	avec toi	0.78

Table 2: Examples of aligned segments found with our methodology. For each candidate, we give its cosine score.

4 Setting up the Annotation Tasks

4.1 Creating Annotation Samples

To generate annotation samples, we applied the algorithm presented in Section 3. First, we filtered tweets based on their similarity scores, removing those with a score below 0.5 under the assumption that such candidates were unlikely to contain PMWEs. This process excluded 10,605 tweets.

Additionally, we removed tweets with a similarity score exceeding 0.99, eliminating another 29,960 tweets, as these were highly likely to contain only MWEs without modifications. Following this filtering, 25,052 tweets remained for annotation.

4.2 Annotation Guidelines

Defining both MWEs and PMWEs from a linguistic and a NLP perspective can be challenging. While linguistic literature does not always agree on all aspects of MWEs (Lamiroy, 2008), PMWEs have been scarcely studied in NLP. For annotation purposes, we adopted the following definitions:

Multiword expression A multiword expression is a fixed sequence of words, either in statistical terms (the words frequently appear next to each other) or in semantic ones (the sequence has a global, non-compositional meaning).

Pun in multiword expression Wordplays or puns created from multiword expressions can be described as multiword expressions that have been unfrozen. To formally identify a wordplay or a pun created from a multiword expression, we must be able to recognise the multiword expression from which it is derived.

Unfreezing Process by which a multiword expression becomes a wordplay or a pun. It involves a formal modification, usually paired with a semantic shift within the multiword expression. This process must not be misjudged for a tense or a number variation, for instance.

We bear in mind that, in the long term, these definitions are intended for non-expert individuals who will learn about these concepts during the annotation process. In addition to these definitions, we give some examples of PMWEs, such as (2), (4) and (6):

1. "*Mangez cinq fruits et légumes par jour*"
   ("eat five fruits and vegetables a day")
2. "*Mangez cinq **riches** et légumes par jour*"
   ("eat five **rich** and vegetables a day")
3. "*Repris de justice*"
   ("convicted")
4. "*Repris de **justesse***"
   ("narrowly recovered")
5. "*C'est le deuxième effet Kisscool*"
   ("it's the second Kisscool effect")
6. "*C'est le deuxième effet **confinement***"
   ("it's the second **lockdown** effect")

(1) becomes (2) (seen at a demonstration in Paris) and (3) becomes (4) (Le Canard Enchaîné, 2017) by word substitution and are well-known MWEs in French. (4) also has a phonetic dimension (ʒys + tɛs VS ʒys + tis). (5) becomes (6) (seen in our corpus) by word substitution as well, but is an older MWE dating from the 80's, so that it may be hard to recognise for some younger speakers. We also introduced true counter-examples found in our corpus, which show variations that do not create a PMWE from a MWE. For instance:

7. "*Max a cassé sa pipe*"
   ("Max kicked the bucket")
8. "*Max avait cassé sa pipe*"
   ("Max kicked the bucket")
9. "*Pierre qui roule n'amasse pas mousse*"
   ("a rolling stone gathers no moss")
10. "*Pierres qui roulent n'amassent pas mousse*"
    ("rolling stones gather no moss")

(8) shows a tense change and (9) a number change. Nevertheless, these 2 examples do not contain any PMWE. They show minor variations of MWEs that mustn't be confused with unfreezing processes, as specified in our PMWE definition.

5 Manual Annotation Task

The annotation task was performed by 3 annotators, A_1, A_2, and A_3, who are also authors of this paper. All had prior experience working with MWEs and PMWEs and had participated in previous annotation tasks. A_3 specializes in linguistics while A_1 and A_2 work in NLP and computer science. The participants were asked to answer two binary questions:

- Does the tweet contain a PMWE ?
- Do you recognize a MWE, unfrozen or not ?

The goal was to directly identify PMWEs without requiring further analysis. After each annotation phase, adjudication sessions were conducted to review the annotations, discuss encountered issues, and resolve disagreements.

5.1 Annotation Phase I: Pilot

Initially, 100 tweets were provided to all three annotators without additional information (such as guidelines or the seed used to fetch them). This sample aimed to assess the difficulty of the annotation task and the annotators' intuition. Krippendorff's (Krippendorff, 2013) α score was 0.19, indicating a significant lack of agreement and highlighting the complexity of identifying PMWEs. An adjudication session followed, where annotators reviewed each tweet and collaboratively established the first set of annotation guidelines.

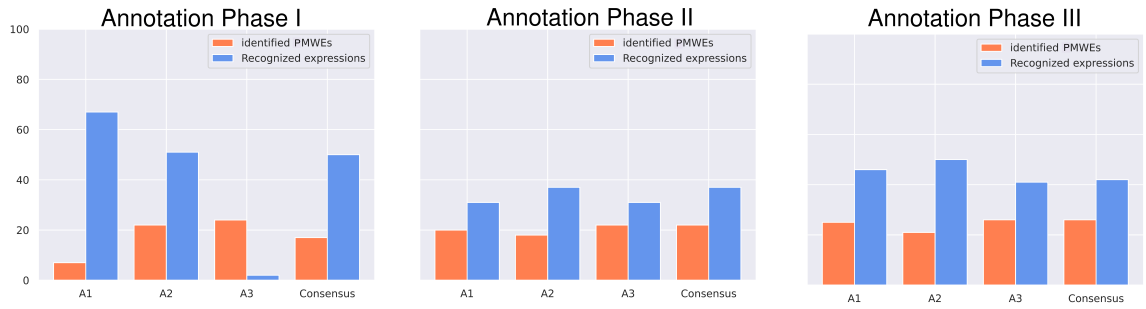

Figure 2: Number of identified PMWEs and recognised MWEs for each annotator and consensus on our three annotation samples.

5.2 Annotation Phase II: First Consolidation

A second set of 100 tweets was provided to the annotators using the newly established guidelines. The resulting Krippendorff's α score improved significantly to 0.77. However, the adjudication session revealed that this sample was easier to annotate due to the high recognizability of PMWEs, leading to fewer disagreements.

5.3 Annotation Phase III: Second Consolidation

A final common sample of 100 tweets was provided. The initial Krippendorff's α score was 0.67, lower than in Phase II but still an improvement over the pilot study.

Discrepancies in annotation strategies emerged: A_1 and A_2 focused on formal changes in MWEs, while A_3 placed greater emphasis on contextual influences. Additionally, A_3 was stricter about variations in quotations and MWEs involving word order changes. Based on these observations, we corrected our annotation guidelines, as explained in Section 6 and each annotator revised its annotations for this sample. The α score for this phase increased to 0.83.

5.4 Annotation Phase IV: Individual Annotations

Beyond the three annotation phases, we proceeded to an individual annotation phase in which each annotator was allocated an additional 100 tweets to annotate.

5.5 Manual Annotation Overview

In total, we annotated 600 tweets. Table 3 shows the frequency of each annotation type across the steps of our annotation process. Of the 600 annotated tweets, 137 (22.83 % of the annotated tweets) were identified as containing a PMWE, whereas 251 (41.83 %) contain only a MWE and 212 (35.33 %) contain nothing. Notably, all identified PMWEs were consistently paired with a recognised MWE. This is expected, as a PMWE should always be linked to an underlying MWE.

Figure 2 presents the number of identified PMWEs and recognised MWEs by each annotator across each annotation phase. We included a consensus column that reflects the final annotations after adjudication. Figure 3 displays the merged confusion matrix for all three annotators.

PMWE	MWE	I	II	III	IV	Total
+	+	17	22	26	72	137
+	-	0	0	0	0	0
-	+	50	37	42	122	251
-	-	33	41	32	106	212
		100	100	100	300	600

Table 3: Frequency of annotations at each step of the manual annotation process : Pilot (I), First consolidation (II), Second consolidation (III) and Individual (IV).

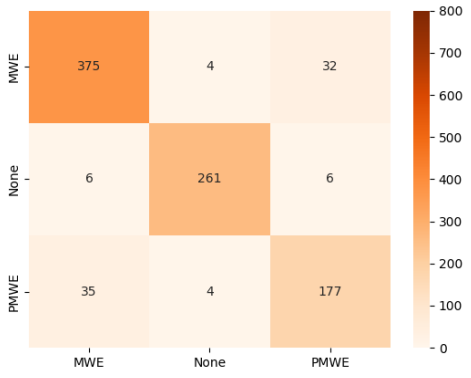

Figure 3: Merged confusion Matrix for the 3 annotators on the 300 tweets they annotated in common.

6 Issues Encountered During the Manual Annotation Task

Throughout the manual annotation process, we identified three major discrepancies between annotator A_3 and the other two annotators: (i) A_3 considered contextual influences more heavily, (ii) applied a stricter approach when annotating MWEs derived from quotations, and (iii) exhibited a different stance on MWEs with word order changes. These differences may stem from A_3's linguistic background, whereas A_1 and A_2 specialise in NLP. Below, we explain how we addressed these discrepancies and refined our annotation guidelines to minimise ambiguity in future PMWE-related annotation tasks.

(i) Contextual influences Although this is a rare scenario, a MWE can unfreeze itself without undergoing a formal modification (Eline and Zhu, 2014). In such cases, only the surrounding context signals the presence of a PMWE. Following the adjudication mentioned in Section 5.3, we decided not to annotate as PMWE any MWE where contextual influences alone reveal a PMWE. This type of PMWE is both infrequent and challenging to identify, introducing significant complexity and inconsistency to the annotation task.

(ii) MWEs corresponding to quotations A_1 and A_2 allowed for minor variations in MWEs originating from well-known quotations. For example, the meme-derived phrase "*Moi je trouve la question elle est vite répondue*" ("I think the question is quickly answered") was frequently truncated to "*La question elle est vite répondue*" ("The question is quickly answered"). While A_3 annotated this as a PMWE, A_1 and A_2 did not. To maintain consistency, we opted for a more flexible approach, permitting slight modifications in MWEs originating from quotations.

(iii) MWEs with word order changes Some PMWEs closely resemble their base MWEs, differing only by slight shifts in word order. The most notable example in our dataset was "*Maurice, tu pousses le bouchon un peu trop loin*" ("Maurice, you're pushing things a little too far"), sometimes reordered as "*Tu pousses le bouchon un peu trop loin, Maurice*" ("You're pushing things a little too far, Maurice"). Since this variation does not appear to involve intentional wordplay, but rather an ignorance of the original quote, we chose not to classify it as a PMWE. However, we encountered a case where the MWE "*Que la force soit avec toi*" ("May the force be with you") became "*avec toi la Force est*"("with you the force is") in a tweet. In this case, the unfreezing process deliberately played with the original word order, so we decided to annotate it as a PMWE.

We also noticed that annotators sometimes repeated the same mistakes from previous annotation phases. To minimise this, we decided to share all consensus annotations among annotators. This way, whenever an annotator encounters a previously discussed case, they can easily refer to our established decision. Moving forward, we plan to leverage our manual annotation findings to develop a GWAP for annotating both MWEs and PMWEs in tweets. This approach will allow us to collect a larger number of annotations efficiently and is presented in the next section.

7 Expanding Annotations with a GWAP

To scale up the annotations of the FRUIT corpus, we developed a participatory science task in the form of a Game With a Purpose (GWAP). This initiative incorporates lessons from our manual annotation task to improve both accuracy and participant engagement.

7.1 Annotation Task Design

Players assume the role of investigators tracking a criminal organisation that manipulates MWEs to conceal hidden messages. Their mission is to identify tweets containing disguised MWEs (PMWEs), following the guidelines established in Section 4. For each tweet, the game highlights a potential MWE and players have to determine (i) if they can identify the indicated MWE and (ii) if this MWE corresponds to a hidden message (i.e. a PMWE). Figure 4 provides a screenshot of the annotation interface.

To encourage engagement, the game features a scoring system and badge collection: players earn points when they annotate a tweet and receive badges when they annotate multiple tweets sharing the same MWE (see Figure 8). Each badge has a design associated with its corresponding MWE. By gamifying this annotation task, we aim to attract different types of players, such as the ones described in Bartle (1996).

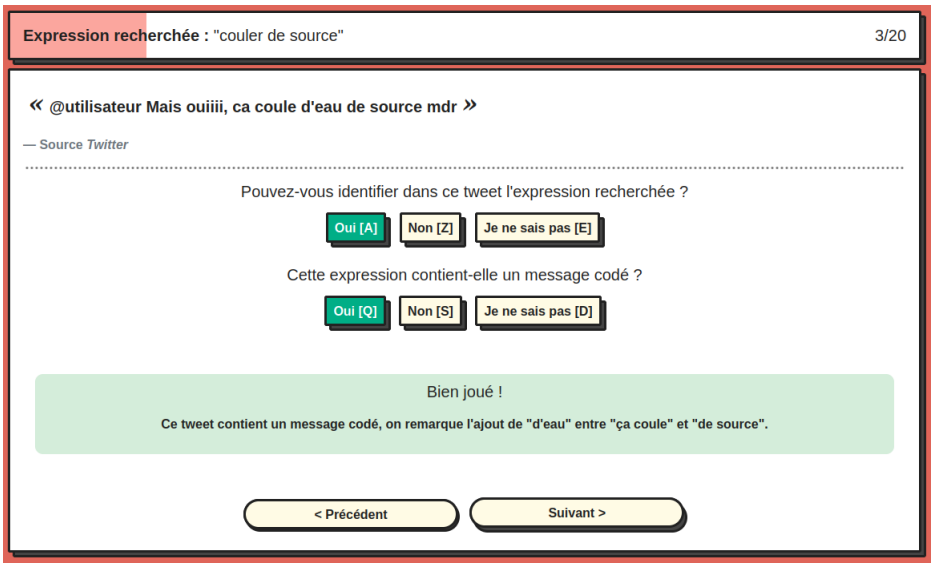

Figure 4: Instance of a tweet to annotate in our GWAP. The upper box contains the indicated MWE, while the lower box contains the tweet and the questions we ask the players to answer. The green box contains the correction given for this tweet.

7.2 Progressive Learning

As shown in Section 6, identifying PMWE and even MWE can be ambiguous. To address this challenge, we incorporated several features into our GWAP to help players gradually learn key concepts related to MWE and PMWEs. Figure 9 illustrate this GWAP annotation process.

Guidelines Players receive a simplified version of the guidelines from Section 4. Prior research shows that clear instructions significantly improve annotation accuracy (Nédellec et al., 2006; Hiebel et al., 2022).

Training set Previous studies suggest that training annotators enhances their performances (Dandapat et al., 2009). To this end, we created a training set of 20 tweets, which players must complete before proceeding to the real annotation task. We selected 20 representative tweets from our previous annotation task, illustrating various MWEs and PMWEs to train the players. After the annotation of each of these tweets, we give feedback and corrections, helping them refine their understanding of the task.

Redundant MWE We dynamically generate random sets for each player, with each set containing up to 20 tweets for annotation. All tweets in a set share the same indicated MWE, allowing players to become more familiar with it and produce more consistent annotations. Once a set is completed, a new one is generated. Players can always revisit previous annotation sets to review or revise their work, fostering continuous learning.

Control Tweets To ensure annotation quality, we randomly distribute 80 of the 600 annotated tweets in Section 5 as control tweets. These tweets have been selected because of their unambiguous annotations. Players receive immediate feedback on these tweets, reinforcing learning and improving consistency. Control tweets can be annotated more than once by a player, allowing us to assess the player's consistency over time.

7.3 Playerbase

As for now, our GWAP has been tested with a limited number of researchers with varying degrees of familiarity with both MWEs and PMWEs. We count 13 players, including A_2, who had not worked on the annotation of PMWEs for over a year at that time. All players speak fluent French and work either in linguistics, computer science or literature. We plan to expand the annotation task available to a wider audience soon.

7.4 Annotation Results

2,206 annotations were made by the 13 players, with an average of 169.7 annotations per player. In total, 931 unique tweets were annotated (1,031 by taking into account training and control tweets). Figure 5 shows the distribution of tweets per number of annotations. We computed an α score of 0.70

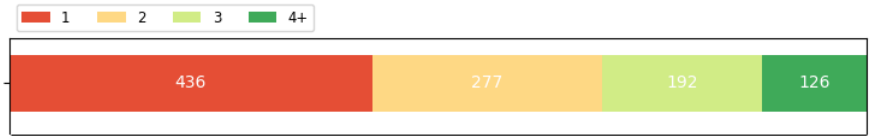

Figure 5: Discrete distribution of tweets per number of annotations for every tweet annotated at least once.

PMWE	MWE	N	R	P	F
+	+	61	92.14	93.14	92.63
+	-	0	/	/	/
-	+	29	85.43	84.61	85.02
-	-	10	76.59	81.81	79.12
Mean			84.72	86.52	85.59

Table 4: Recall (R), precision (P) and F-score (F) obtained by comparing annotations made by the players with annotations made by the experts for each possible annotation made.

by taking into account every tweet which were annotated more than once (595 tweets, training and control tweets included). We compared our crowdsourced annotations on the training and control tweets with the annotations made by the experts in Section 5. All the 80 control tweets and the 20 training tweets were annotated more than once, therefore, we include them all in this comparison. Table 4 summarises the results we obtained for each annotation category.

We observe that the mean F-score is high (85.59), indicating a high level of agreement between players and experts. Surprisingly, PMWE identification has a better F-score than MWE recognition (92.63 against 85.02). This can likely be attributed to the fact that our guidelines are more focused on PMWEs. No annotator has identified a PMWE without recognising a MWE, which is why we do not report metrics for this particular scenario. Table 7, 8 and 9 in the Appendix show the 100 tweets (control + training) given to our players.

8 Discussion

In this paper, we introduced the FRUIT corpus, containing 60,617 tweets among which 1,531 have been manually annotated through (i) an expert review and (ii) a GWAP. The results of the manual annotation task show that both MWE and PMWE identification tasks are challenging, even for experts with substantial experience in these two notions. We argue that the low inter-annotator score of 0.19 obtained during our pilot annotation (Section 5.1), alongside the discussion presented in Section 5.3, may be attributed to differences in the interpretation of MWEs and PMWEs between NLP experts and linguistics experts. Despite these challenges, by developing clearer guidelines and organising adjudication sessions, we improved our understanding of both MWEs and PMWEs, which likely contributed to an increase in our inter-annotator score to 0.83.

The GWAP demonstrates that it is possible to teach non-expert individuals how to recognise and identify both MWEs and PMWEs. To achieve this, we leveraged the guidelines developed during the manual annotation task. We also allowed our players to improve their understanding of the key notions through progressive learning (Section 7.2). The results exhibit a high level of agreement between players, with an inter-annotator score of 0.70. Furthermore, we unveil that our players tend to agree with the reference annotation made by our three experts, with an observed mean F-score of 85.59 for every type of annotation (92.63 for PMWE identification).

This result might be influenced by the fact that our players are primarily from the research area, and some of them having already basic knowledge on MWEs and occasionally PMWEs. Despite this potential bias, the insights obtained from this annotation task will inform future improvements to the GWAP and the annotation process.

Looking ahead, we intend to continue annotating the FRUIT corpus through the GWAP presented here. In particular, we want to make this GWAP available to a wider non-expert audience so that we can observe the quality of our progressive learning. We also plan to create a second annotation task, whose goal will be to annotate found PMWEs at different levels.

We plan to assemble a multilingual dataset containing MWEs and PMWEs from films and article titles (media and scientific). Such a dataset could help us analyse differences in PMWE construction across languages. This future work could benefit from a participatory annotation task, such as the one described here.

Ethical Considerations

We have ensured that our annotators remain anonymous. To sign up for GWAP, we only ask for a username and password, without collecting any additional data. We have also anonymised every tweet in the FRUIT corpus. Finally, we inform players of the potential presence of offensive content in tweets (violence, hatred, inappropriate content, etc.). If a player identifies an offensive tweet, we invite them to contact us so that we can deal with it.

Acknowledgments

We would like to thank Karën Fort for her suggestions, including the addition of collectable badges, and her help while designing the annotation by sets used in our GWAP. Finally, we thank all the players who participated in our GWAP.

References

Timothy Baldwin and Su Nam Kim. 2010. Multiword Expressions. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing*, 2 edition. Chapman and Hall/CRC.

Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit muds. *The Journal of Virtual Environments, 1*.

Julien Bezançon and Gaël Lejeune. 2023. Reconnaissance de défigements dans des tweets en français par des mesures de similarité sur des alignements textuels. In *30e Conférence sur le Traitement Automatique des Langues Naturelles, TALN*, pages 56–67, Paris, France. ATALA.

Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner, and Amalia Todirascu. 2017. Survey: Multiword Expression Processing: A Survey. *Computational Linguistics*, 43(4):837–892. Place: Cambridge, MA Publisher: MIT Press.

Christophe Cusimano. 2015. Figement de séquences défigées. *Pratiques*, (159-160):69 78.

Sandipan Dandapat, Priyanka Biswas, Monojit Choudhury, and Kalika Bali. 2009. Complex linguistic annotation – no easy way out! a case from Bangla and Hindi POS labeling tasks. In *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, pages 10–18, Suntec, Singapore. Association for Computational Linguistics.

Joël Eline and Lichao Zhu. 2014. Défigement et inférence - cas d'études du Canard enchaîné. *SHS Web of Conferences*, 8:681 695. 1 citations (Crossref) [2023-11-09] 0 citations (Semantic Scholar/DOI) [2022-11-14].

Liana Ermakova, Anne-Gwenn Bosser, Tristan Miller, Victor Preciado, Grigori Sidorov, and Adam Jatowt. 2024. *Overview of the CLEF 2024 JOKER Track: Automatic Humour Analysis*, pages 165–182.

Liana Ermakova, Tristan Miller, Julien Boccou, Albin Digue, Aurianne Damoy, and Paul Campen. 2022. Overview of the clef 2022 joker task 2: translate wordplay in named entities. *Proceedings of the Working Notes of CLEF*, pages 1666–1680.

Liana Ermakova, Tristan Miller, Anne-Gwenn Bosser, Victor Manuel Palma Preciado, Grigori Sidorov, and Adam Jatowt. 2023. Overview of joker–clef-2023 track on automatic wordplay analysis. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 397–415. Springer.

Karën Fort. 2022. *Myriadisation et éthique pour le traitement automatique des langues*. Accreditation to supervise research, ED n°77 : Informatique - Automatique - Électronique - Électrotechnique - Mathématiques de Lorraine (IAEM-Lorraine).

Karën Fort, Gilles Adda, and K. Bretonnel Cohen. 2011. Last words: Amazon Mechanical Turk: Gold mine or coal mine? *Computational Linguistics*, 37(2):413–420.

Karën Fort, Bruno Guillaume, Matthieu Constant, Nicolas Lefèbvre, and Yann-Alan Pilatte. 2018. "Fingers in the Nose": Evaluating Speakers' Identification of Multi-Word Expressions Using a Slightly Gamified Crowdsourcing Platform. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 207–213, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Karën Fort, Bruno Guillaume, Yann-Alan Pilatte, Mathieu Constant, and Nicolas Lefèbvre. 2020. Rigor Mortis: Annotating MWEs with a Gamified Platform. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4395–4401, Marseille, France. European Language Resources Association.

Kristina Geeraert, R. Harald Baayen, and John Newman. 2018. *"Spilling the bag" on idiomatic variation*, pages 1–33. Number 2 in Phraseology and Multiword Expressions. Language Science Press.

Gerda Haßler and Christiane Hümmer. 2005. Figement et défigement polylexical : l'effet des modifications dans des locutions figées. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (53):103–119.

Nicolas Hiebel, Olivier Ferret, Karën Fort, and Aurélie Névéol. 2022. CLISTER : A corpus for semantic textual similarity in French clinical narratives. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4306–4315, Marseille, France. European Language Resources Association.

Nicolas Hiebel, Bertrand Remy, Bruno Guillaume, Olivier Ferret, Aurélie Névéol, and Karen Fort. 2024. Hostomytho: A GWAP for synthetic clinical texts evaluation and annotation. In *Proceedings of the 10th Workshop on Games and Natural Language Processing @ LREC-COLING 2024*, pages 14–20, Torino, Italia. ELRA and ICCL.

Klaus Krippendorff. 2013. *Content Analysis: An Introduction to Its Methodology*. SAGE.

Cvetana Krstev and Agata Savary. 2017. Games on Multiword Expressions for Community Building. *Infotheca*, 17(2):7–25.

Mathieu Lafourcade. 2007. Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07: 7th International Symposium on Natural Language Processing*, page 7, Pattaya, Chonburi, Thailand.

Mathieu Lafourcade, Alain Joubert, and Nathalie Le Brun. 2015. *Games with a Purpose (GWAPS)*. John Wiley & Sons.

Béatrice Lamiroy. 2008. Le figement: à la recherche d'une définition. *ZFSL, Zeitschrift für französische Sprache und Literatur*, 36:85–99.

Aude Lecler. 2006. Le défigement : un nouvel indicateur des marques du figement ? *Cahiers de praxématique*, (46).

Chris Madge, Richard Bartle, Jon Chamberlain, Udo Kruschwitz, and Massimo Poesio. 2019. Making text annotation fun with a clicker game. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, FDG '19, New York, NY, USA. Association for Computing Machinery.

Salah Mejri. 2013. Figement et défigement : problématique théorique. *Pratiques. Linguistique, littérature, didactique*, (159-160):79–97. 3 citations (Crossref) [2023-11-09] 2 citations (Semantic Scholar/DOI) [2022-11-15] Number: 159-160 Publisher: Association CRESEF.

Tristan Miller, Christian Hempelmann, and Iryna Gurevych. 2017. SemEval-2017 task 7: Detection and interpretation of English puns. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 58–68, Vancouver, Canada. Association for Computational Linguistics.

Claire Nédellec, Philippe Bessières, Robert R. Bossy, Alain Kotoujansky, and Alain-Pierre Manine. 2006. Annotation guidelines for machine learning-based named entity recognition in microbiology. In *Proceeding of Data and Text Mining for Integrative Biology Workshop 17. European Conference on Machine Learning 10. European Conference on Principles and Practice of Knowledge Discovery in Databases*, Workshop on data and text mining for integrative biology, Berlin, Germany. Springer - Verlag. ON LINE.

Geoffrey Nunberg, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. *Language*, 70:491–538.

Alan Scott Partington. 2009. A linguistic account of wordplay: The lexical grammar of punning. *Journal of Pragmatics*, 41(9):1794–1809.

Carlos Ramisch, Silvio Cordeiro, Leonardo Zilio, Marco Idiart, and Aline Villavicencio. 2016. How Naked is the Naked Truth? A Multilingual Lexicon of Nominal Compound Compositionality. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 156–161, Berlin, Germany. Association for Computational Linguistics.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In *Computational Linguistics and Intelligent Text Processing*, Lecture Notes in Computer Science, pages 1–15, Berlin, Heidelberg. Springer.

Agata Savary, Silvio Cordeiro, and Carlos Ramisch. 2019. Without lexicons, multiword expression identification will never fly: A position statement. In *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, pages 79–91, Florence, Italy. Association for Computational Linguistics.

Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova, and Antoine Doucet. 2017. The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 31–47, Valencia, Spain. Association for Computational Linguistics.

Alessandro Valitutti, Hannu Toivonen, Antoine Doucet, and Jukka M Toivanen. 2013. "let everything turn well in your wife": generation of adult humor using lexical constraints. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 243–248.

A Appendix

Table 5 shows the number of tweets of the FRUIT corpus filtered at each step. Figure 6 shows the confusion matrices obtained at the end of our manual annotation task. We discuss several aspects regarding our methodology for building the FRUIT corpus in Section A.1. In Section A.2, we further describe our GWAP.

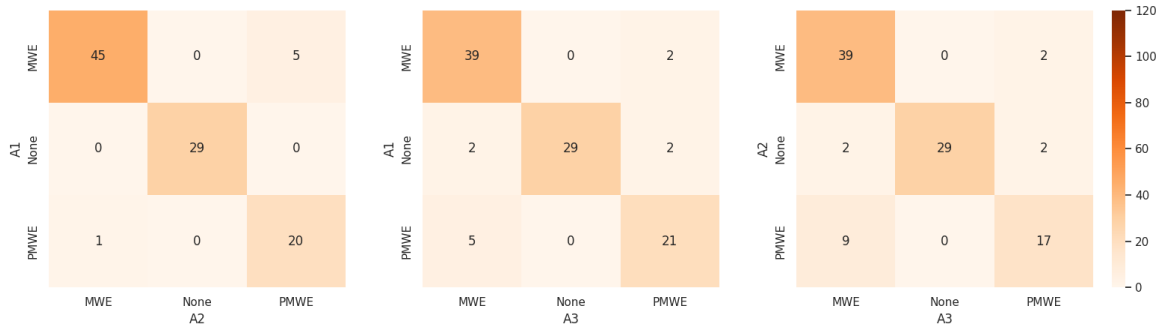

Figure 6: Confusion matrix for each annotation pair for the 300 tweets annotated in common by the 3 annotators.

	Initial	**< 50 %**	**Dup.**	**By seeds**
Filtered	/	3,268,394	15,381	20,244
Total	3,369,636	101,242	85,861	60,617

Table 5: Statistics on each filtering step. **< 50 %** corresponds to the number of tweets with less than 50 % of the words of their *seed*, **Dup.** to filtered duplicate tweets and **By seeds** to tweets filtered according to their *seed*.

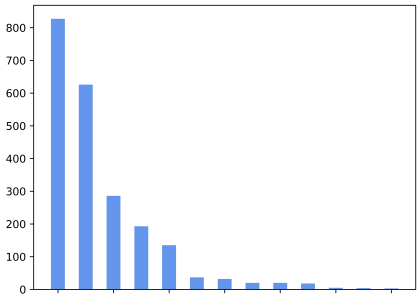

Figure 7: Number of annotations produced by each player.

A.1 Corpus Building Details

Each query made on Twitter consisted of one of our MWEs. We queried Twitter daily, issuing one query per MWE. Among the returned tweets, we only retained those that contained more than half of the words in the corresponding MWE, filtering out the rest.

These MWEs were primarily selected for their conventional nature, which mean that they must remain recognisable to a broad audience. We adopt a broad definition of MWE, encompassing verbal MWEs, phrasemes, collocations, idioms, and even citations, especially well-known ones, as they tend to be conventionalised. For example, we consider a citation such as "*travailler plus pour gagner plus*" a MWE because (i) it is conventionalised, and (ii) it carries an additional meaning, making it somewhat non-compositional. This particular citation, used by Nicolas Sarkozy in 2007, is now often referenced satirically as a symbol of capitalism.

To compute similarity scores, we vectorised each candidate and seed expression using the TFIDFVECTORIZER feature from the SCIKIT-LEARN library. We used word bigrams and trigrams. This process was repeated across multiple linguistic representation layers obtained with the SPACY library, incorporating POS tags and lemmas in addition to tokens.

A.2 Annotation Tasks Details

We take into account the fact that the FRUIT corpus is imbalanced (86.51 % of the tweets were found with the top 10 first *seeds*) when creating our annotation samples. For the manual annotation task, each sample was created using a maximum of 5 tweets related to the same *seed* to ensure diversity. For our GWAP, we limited to 500 the maximum number of tweets for a *seed*, randomly selecting 500 tweets if a *seed* has more than this number. We plan to add more tweets over time.

Figure 9 summarises the annotation process we implemented in our GWAP. Figure 7 shows the number of annotations made by each player, while Figure 8 shows the top four players in our ranking system. More annotations were made during the redaction of this paper, which is why the scores shown here are higher than the number of annotated tweets we indicate. Table 6 contains every tweet used for the training phase of our GWAP, alongside with the consensus annotation made during the manual annotation task. We also show our control tweets in Table 7, Table 8 and Table 9.

Username	Score	Badges
Michel	1132	
ChatGBouté	711	
maxx_leh	681	
Poutpout	298	

Figure 8: Top 4 players in our ranking.

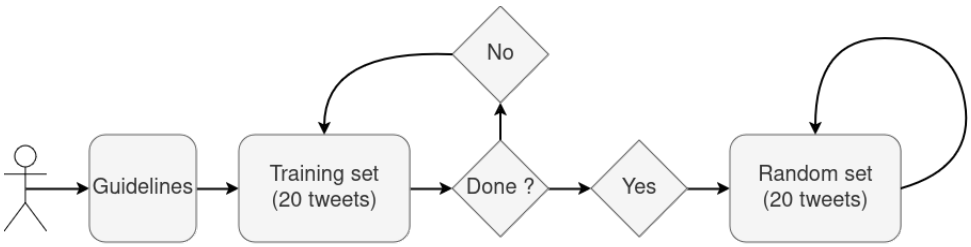

Figure 9: Summary of the annotation process we implemented in our GWAP.

Tweet	MWE	PMWE
« Pourquoi ils n'ont pas de programme ? Parce que le programme de Giscard et de Macron c'est le même : **travailler plus pour même plus gagner plus** et réduire les impôts des riches. Ca fait 50 ans qu'il existe ce programme, vaut mieux pas qu'il l'énonce ! » https://t.co/m28301zbZJ	+	+
@utilisateur @utilisateur **Travailler plus pour gagner moins** !!!	+	+
@utilisateur Sans oublier qu'il s'agit de salariés ayant un niveau de vie "confortable" (euphémisme) sans difficultés à boucler leurs fins de mois, donc absolument pas motivés à "**travailler plus pour gagner autant**".	+	+
@utilisateur **C'est le deuxième effet covid**	+	+
@utilisateur_Danaos C'était peut être pas son intention mais c'est le résultat. Le deuxième effet étant une réserve de voix au second tour ...	-	-
@utilisateur Mais ouiiii, **ca coule d'eau de source mdr**	+	+
@utilisateur ça coule de source après donc bon	+	-
Tant qu'il y a de l'amour il y a de la vie! https://t.co/AU0mrWeGJa	+	+
J'aime mon pays France mondial j'aime la planète Terre l'eau le vent le gel le froid le soleil la lune la nuit l'hiver l'été l'automne et le printemps quand je vois tout cela **tant qu'il y a la vie il y a de l'amour il y a de l'espoir** j'aime la planète Terre	+	+
Travailler pour plus tard la gâter c'est mon objectif	-	-
@utilisateur **Casse toi en Espagne pauvre con**	+	+
@utilisateur_pic @utilisateur Et pour compléter, tous les profs du secondaire ne l'ont pas mais tous font de l'orientation etc. Alors oui c'est injuste. Mais ce que tu décris ce sont des missions liées à des primes. **Travailler plus pour gagner à peine plus** n'était pas le sujet initial. Bonne journée.	+	+
La question, elle est vite et parfaitement répondue ce samedi par Eric Neuhoff (qui a regardé les #Cesar2021 jusqu'au bout, lui...) sur le site du @utilisateur_Figaro. C'est oui. https://t.co/10cBkZXMFe	+	+
@utilisateur C'était un beau pays la France. Mais elle n'est plus. Plus aucune valeur, plus rien. Le combat de certains derniers irréductibles est vain, perdu d'avance. **Égalité , fraternité , liberté=confiné**.	+	+
@utilisateur__anton Nn toi en ce moment tu fais l'aigri, **ma France tu l'aimes ou tu l'as quittes** fin .	+	+
Mohamed SALAH **que la force de l'Égypte Antique soit avec toi** Ouvre le chapitre vengeance face au RÉAL MADRID https://t.co/Pp9trT29WF	+	+
@utilisateur **Que la Force (du Droit) soit avec toi** alors ! En souhaitant qu'en plus vous trouviez un meilleur appart' !	+	+
@utilisateur @utilisateur @utilisateur Mais bon faire autrement ce serait discriminatoire... . Le patriarcat... C'est fini ou pas ? A un moment faut prendre position ! **Le beurre l'argent du beurre et les glawis du crémier**... Ca va 5mn!	+	+
@utilisateur @utilisateur @utilisateur_ alors là Maurice tu pousses le bouchon un peu trop loin	+	-
@utilisateur Ce dossier survient au plus mauvais moment pour Emmanuel Macron dans la mesure où celui-ci fournit des armes de destruction massive à son(a) futur(e) adversaire du second tour. L'épilogue de ce scrutin présidentiel devient désormais indécis.	+	-

Table 6: Training tweets given to our players, with the consensus annotations from the manual annotation phase. We highlight PMWEs in bold and underline MWEs.

Tweet	MWE	PMWE
Tu préfères être suivi(e) par Christine Lagarde dans ton sommeil comme une personne de basse classe sociale ou bien épiler des maîtres chiens à chaque fois que tu rencontres une nouvelle personne ? Moi je pense la question elle est vite répondue. Bisous.	+	-
@utilisateur_Opin @utilisateur L'ourse est morte à cause d'un type venu chez elle, armé jusqu'aux dents et avec l'intention de tuer. Pour moi la question est vite répondue !	+	-
TRAVAILLER PLUS POUR GAGNER MOINS L'accord pour la modernisation des ressources humaines de la police nationale 2022/2027 signé par les syndicats « maison » est historique… POUR LA 1RE FOIS ILS ONT ACTÉ LA FUTURE BAISSE DE SALAIRE ! Lire en ligne https://t.co/P3BMuAp1Xp https://t.co/jz2YdfDZ2n	+	+
Nous ne sommes pas les seuls êtres vivants sur terre. Quand nous gérons mal nos déchets, c'est les animaux qui en souffrent ! #StopPollution https://t.co/Jq2oqHTuQo	-	-
@utilisateur Et nous ne sommes pas les seuls, j'en suis convaincu...	-	-
@utilisateur @utilisateur @utilisateur @utilisateur La thrombose c'est la protéine Spike en revanche. L'oxyde de graphène c'est pour plus tard, c'est le deuxième effet kiss cool. Bon rétablissement à lui	+	-
@utilisateur Ça coule de source hehe	+	-
@utilisateur Çà coule de source	+	-
@utilisateur Comment faire pour discréditer une personne, et bien tous les coups sont permis. Pauvre France, c'est ça qu'on appelle liberté, égalité, fraternité. Vive Reconquête, vive Eric Zemmour et vive la France	+	-
@utilisateur @utilisateur_Danaos @utilisateur On se connait ? Non. Alors faut s'en tenir à ce que vous connaissez. Dès le moment où on avoue que tous les coups sont permis car "c'est la campagne" vous discréditez le politique. Un cirque. Pas plus. Et tout ce qui sera dit pourra être remis en cause à travers ce prisme.	+	-
@utilisateur_morel Ces gens utilisent un vocabulaire complexe qui demande d'avoir étudié et pratiqué. Mais là l'éducation qu'ils ont reçue ne semble pas soutenue par de l'intelligence. On a donné de la confiture aux cochons…	+	-
@utilisateur_Desouche @utilisateur Bravo pour l'initiative de toute façon c'est donner de la confiture à des cochons quand on voit ce qu'ils font des quartiers, dommage car il y'a certainement des gens bien qui vont en pâtir à cause de ces raclures	+	-
@utilisateur_ Non tkt ça va aller on y croit que la force soit avec toi	+	-
@utilisateur_Ringo Ouai enfin tu comprends rien visiblement x) Pas grave, bonne journée mon brave et que la force soit avec toi	+	-
Le plus grand chagrin d'amour c'est quand la mort s'en mêle. Tant qu'il y a la vie, il y a de l'espoir. As long as you live, fight for what you love	+	-
Je ne sais pas ce qu'il reste de ces 3 mots : Liberté, égalité et fraternité !	+	-
@utilisateur **Travailler moins pour gagner plus** donc voilà votre solution ? heureusement vous serez jamais au pouvoir	+	+
@utilisateur @utilisateur Il m'est arrivé la même chose ; nous ne sommes pas les seuls, malheureusement… de plus en plus de censure !!!	-	-
Ça va coulé de source #adp2020 https://t.co/nRga33whi6	+	-
@_NdRoussel @utilisateur_steiger Ça fait grave penser à "la France tu l'aimes ou tu la quitte" de Sarko. Y'a des moods chelou au PCF en ce moment.	+	-
Allez on inaugure ces perles en beauté ! **Que la force d'Apoula Edel soit avec toi** champion https://t.co/xkpU2JOAtm	+	+
@utilisateur_blond Regard au delà doit porter, **que la force de découverte avec toi, soit**	+	+
@utilisateur aller ryzeuh **que la force du cookie monster soit avec toi**	+	+
Rien ne les obligent a travailler a la SNCF. Il y a plein de jobs ouverts pour lequels les horaires sont plus souples. Vous voulez **la vache, le lait, le beurre et l'argent de la cremiere**. Ne plus céder aux methodes marxistes de la CGT, c'est la seule solution. https://t.co/iu01c3VOFG	+	+
@utilisateur_BLITZERS Ils font le tour du monde en ce moment ou quoi ? J'ai raté un épisode ?	-	-
@utilisateur_Lol Maurice tu pousses le bouchon un peu trop loin	+	-
@utilisateur Là j'avoue tu pousses le bouchon un petit peu trop loin Maurice!	+	-
@utilisateur **Jean tu pousses le bouchon un peu trop loin et T es pas Maurice**	+	+
@utilisateur **C'est le deuxième effet coupe du monde**	+	+
@utilisateur C'est le deuxième effet kiss cool	+	-

Table 7: Control tweets given to our players, with the consensus annotations from the manual annotation phase. We highlight PMWEs in bold and underline MWEs.

Tweet	MWE	PMWE
@utilisateur **C'est le deuxième effet qui se coule** de la politique de Biden: balkaniser l'UE quand les victimes de l'ultra libéralisme vont commencer à être déstabilisés donc livrés aux mafias!	+	+
Il y a bien longtemps, dans une galaxie très lointaine... @utilisateur_LiT_Sand	+	-
Il y a bien longtemps, dans une galaxie très lointaine...	+	-
La question est vite répondue : le public vote à l'unanimité pour @utilisateur. https://t.co/c33004Zyuk	+	-
@utilisateur Pour **gagner plus et travailler moins** pardi. Ça ne va pas aider les gens à trouver facilement un médecin.	+	+
@utilisateur **C'est le deuxième effet grumpy** lol	+	+
C'est plus le deuxième effet kisscool, **c'est le deuxième effet médiapart**: y a toujours une deuxième révélation après la première pour enfoncer le clou https://t.co/4ZAHfWRQEK	+	+
La vie n'essaie pas de la prévoir La vie c'est la pluie, le beau temps C'est une larme, des souvenirs Des espoirs de l'amour C'est un sourire a tes lèvres. https://t.co/zqGRxkrX8m	-	-
@utilisateur **Que la force du #digital soit avec toi** ! https://t.co/4BOoO2Qlgj	+	+
@_clemparker_ **Que la fibre.. euuuh la force soit avec toi**..! Et là-bas, tu auras un nouveau chez toi.	+	+
@utilisateur Euh je ne trouve pas c'est un tweet qui reflète malheureusement une triste réalité. Mais bon Zazou tu dois faire partie de ces gens qui pensent que l'**on peut et doit accueillir toute la misère du monde**. J'ai hâte qu'ils frappent à ta porte	+	+
En "douce France de l'omerta", n'aurais été victime d'agressions crapuleuses, frappes répétées, LGBTI Phobies caractérisées homophobes, d'humiliation, d'harcèlement & bénéficié d'aucune hospitalisation! **Liberté égalité dignité fraternité justice**?! > https://t.co/Q6C6cb5ihd https://t.co/2tAz9NA6Kt	+	+
@utilisateur_C_O_N_S **Tu pousses le bouchon un peu trop loin Farid pour ne pas t'appeler Maurice** grrrrrr	+	+
@utilisateur Le mec a peur que les grands méchants patrons utilisent le pied dans la porte pour faire travailler les pauvres employés plus, mais diminuer les salaires unilatéralement par le saccage monétaire c'est OK	-	-
@utilisateur @utilisateur_liberal On est en train de toujours s'occuper à travailler plus pour l'occupation et l'agitation de nos démarches au niveau de. Point.	-	-
@utilisateur **c'est le deuxième effet du décolleté d'hier**? (soignes toi bien)	+	+
@utilisateur @utilisateur @utilisateur @utilisateur_ Bonne chance ! **Que la force d'Eren soit avec toi**	+	+
Bon vent @utilisateur. **Que la force du panda soit avec toi**. https://t.co/AAAPfSJTKd	+	+
@utilisateur Merci pour l'info et **que la Force de guérir soit avec toi** !	+	+
@utilisateur_ghostz @utilisateur **À défaut de la Force, que la chance soit avec toi** @utilisateur Comme on dit: Fingers crossed	+	+
@utilisateur_canna Looooool **que la force de la weed soit avec toi** !	+	+
@utilisateur_ "Alors, tu préfères **le beurre, l'argent du beurre ou le cul de la crémière** ? Pour moi, la question elle est vite répondue" https://t.co/CgyZcXgKMj	+	+
@utilisateur On ne sait pas mais **peut on encore se permettre d'accueillir toute la misère du monde ?**	+	+
@utilisateur Hé **Maurice macron tu pousses le bouchon un peu trop loin** tout va te péter à la G.... (en 6 lettres) Achtung achtung ... Pour que cette folie s'arrête je sais ce qu'il faut faire mais j'vous l'dirai pas ou du moins pas tout suite ! Tout arrive à celui qui sait attendre ...	+	+
J'ai encore lu que Macron veut «augmenter les profs qui travailleront plus». Ça a été dit 100 fois mais rappelons quand même que ça n'a aucun sens. Une augmentation c'est **gagner plus sans travailler plus**. **Gagner plus en travaillant plus** c'est juste normal.	+	+
@utilisateur, Conseiller Regional @utilisateur, soutient les salarié•es de #BREGAMS en lutte contre un Accord de Performance Collective (APC) qui les fait **travailler plus pour gagner beaucoup moins**! https://t.co/19ANl3MAZo https://t.co/kfalDVFJRE	+	+
@utilisateur @utilisateur @utilisateur Ça nous coûtera notre maison et tout nos biens, même nos enfants et notre corps, quand il faudra payer l'addition de l'argent magique dans quelques années. C'est le deuxième effet kisscool, le plan pour installer une société comme en Chine et justifier l'injustifiable.	+	-

Table 8: Control tweets given to our players, with the consensus annotations from the manual annotation phase. We highlight PMWEs in bold and underline MWEs.

Tweet	MWE	PMWE
@utilisateur **C'est le deuxième effet du coup de coeur vaccin**	+	+
@utilisateur_Stream **Que la force du requin soit avec toi**	+	+
@utilisateur **Que la force de l'amour soit avec toi** pour vaincre cette saloperie !	+	+
@utilisateur **Que la (Tri)force soit avec toi** !	+	+
@utilisateur_philippot Gros con qui se la joue plus français que tout le monde. La devise c'est liberté, égalité, fraternité. Le reste n'est pas français. Traître.	-	-
@utilisateur_dufour Ils se vengent de votre position. Assumez	-	-
@utilisateur_trading @utilisateur @utilisateur @utilisateur Vous n'avez toujours pas compris que votre position est nauséabonde parce qu'elle se défile/cache derrière une notion juridique qui n'a pas de sens d'être utilisée, au lieu de tout simplement assumer le fait de dire "je ne crois pas les victimes". Dites le, allez, assumez un peu.	-	-
jusqu'ici tout va bien	+	-
@utilisateur Et ces gens qui soutiennent Macron vont eux-aussi impactés par cette **politique de destruction massive** du modèle économique et social issu du CNR, des idéaux républicains etc.	+	+
@utilisateur_1ere Mais qu'attendent donc la #NUPES et toutes ces **ONG de destruction massive** pour enfin revendiquer le droit de cuissage, ou une dotation de quelques vierges, pour tout migrant illégal qui arriverait en #France? Un accueil digne pour ces pauvres gens, serait la moindre des choses.	+	+
Terrible !Ces gens là sont nos pires ennemis : cette oligarchie mondialiste qui se trouve dans tous ces pays qui ont participé à cette mascarade criminelle. Ceux-ci ont utilisé les pays à leur solde comme **instrument de destruction massive** contre les peuples d'y trouvant:tromperie https://t.co/7G5U6x8QXP	+	+
NOUVEAUTE / Sinaïve, Dasein EP (Buddy Records) / Reprise Party (Langue Pendue) **Il y a bien longtemps, dans une très petite galaxie fort lointaine**, nous nous battions au sujet de l'usage de la langue française dans un contexte noisy pop. Par @utilisateur https://t.co/e4QHjnWGqS	+	+
Rappel : la seule "nouvelle mission" qui intéresse Macron sera les remplacements bouche-trous. Soit, compte tenu de l'inflation, **travailler plus pour gagner aussi peu qu'avant**. Et ça passera, parce que la profession est désormais dépolitisée. #Cassandre https://t.co/GSwCFTlZgX	+	+
@utilisateur_morel Cette photo montre aussi, qu'hormis la petite foule de "journalistes" qui se pressent autour de la Raclure néo nazie, la salle, qu'on aperçoit à l'arrière plan, est déserte Ça, **c'est le deuxième effet "grand angle"**	+	+
@utilisateur_man_one **Le futur est déjà derrière nous**	+	+
@utilisateur_delb Putain mais j'ai honte pour lui... **À genoux en rampant devant les racistes pseudo-damnés de la terre**	+	+
@utilisateur Mes excuses et courage ! **Que la force des dieux soit avec toi** toujours en ta faveur ! https://t.co/fFDyUl75Sc	+	+
Go mon #Bilou ! **Que la force du champs de lin soit avec toi** ! @utilisateur @utilisateur @utilisateur @utilisateur https://t.co/LONjqVzRRi	+	+
On ne peut pas et accueillir toute la misère du monde en prendre soin à grands coups de milliards et s'occuper de nos bébés placés qui EUX sont notre futur. L'État a clairement fait son choix! https://t.co/bbDFXivU26	+	-
@utilisateur @utilisateur @utilisateur @utilisateur Ah oui si on est contre une immigration non contrôlée on est raciste. Je m'en fout de la couleur de peau ou de la provenance des immigrés. Ce que je souhaite c'est préserver notre mode et niveau de vie. **On ne peut pas et ne veut pas accueillir toute la misère du monde**!	+	+
@utilisateur **Président momo "Maurice" tu pousses le bouchon un peu trop loin** et Il n'y a pas de musulmans modérés.	+	+
@utilisateur Qui a été formé à Bordeaux, joue en principauté et sera le futur joueur du PSG ? La question elle est vite repondue	+	-
Il y aura deux grands choix de société en 2022 : **travailler plus pour gagner pareil** ou **travailler moins pour gagner pareil**. Les innombrables candidats se répartissent dans ces deux catégories. #Presidentielle2022 https://t.co/28M1LA1WnA	+	+

Table 9: Control tweets given to our players, with the consensus annotations from the manual annotation phase. We highlight PMWEs in bold and underline MWEs.