

Cheese it up: CamemBERT Outperforms Large Language Models for Identification of French Multi-word Expressions

Sergei Bagdasarov and Diego Alves and Elke Teich

Saarland University, Germany

sergeiba@lst.uni-saarland.de, diego.alves@uni-saarland.de, e.teich@mx.uni-saarland.de

Abstract

In recent years, language models, both encoder-only and generative, have been applied to a variety of downstream NLP tasks, including sequence labeling tasks like automatic multi-word expression identification (MWEI). Multiple studies show that, in general, fine-tuned encoder-only models like BERT tend to outperform pretrained generative LLMs on downstream tasks (Arzideh et al., 2025; Ochoa et al., 2025; Bucher and Martini, 2024; Sebők et al., 2025). However, such comparisons are sparse for MWEI, in particular for French, in part due to the lack of comprehensive gold-standard datasets. In this study, we address this research gap by comparing CamemBERT with gpt-oss and Qwen3 for MWEI, using the French subcorpus of the newly released PARSEME dataset. CamemBERT outperforms both LLMs by large margins in precision, recall, and F1. We complement this numerical evaluation with a qualitative analysis of prediction errors.

1 Introduction

Multi-word expressions (MWEs) are prefabricated sequences of words that tend to be stored and processed as whole units in memory, rather than composed online from their individual components (e.g. *briser la glace* (break the ice), *lors de* (at the time of), *poser [une] question* (ask [a] question) (Siyanova-Chanturia, 2013; Siyanova-Chanturia et al., 2017). MWEs are ubiquitous in language and are widely attested in different text types, making automatic MWE identification (MWEI) a crucial task in many natural language processing (NLP) applications. At the same time, MWEI remains a challenging problem due to the structural diversity, variability, as well as semantic and syntactic idiosyncrasies of MWEs.

In recent years, transformer-based encoder-only language models have achieved strong performance across a wide range of downstream NLP tasks, including sequence labeling tasks like MWEI (Bello

et al., 2023; Garrido-Merchan et al., 2023; Bui and Savary, 2024; Labusch et al., 2019). In parallel, generative large language models (LLMs) have gained increasing attention since 2022 and are now widely used by scholars for diverse NLP tasks, often in zero-shot or few-shot settings (Gilardi et al., 2023; Törnberg, 2023).

Despite this progress, systematic comparisons between fine-tuned pretrained models and generative LLMs for MWEI remain limited, particularly for languages other than English. This paper addresses this gap by comparing a fine-tuned CamemBERT model (Martin et al., 2019) with two open-source LLMs, gpt-oss-20b (OpenAI, 2025) and Qwen3-32B-AWQ (QwenTeam, 2025), providing a focused evaluation in the context of automatic identification of French MWEs.

2 Related Work

Since the introduction of transformers technology in 2017, pretrained encoder-only models like BERT have proven to be powerful and versatile tools that found their applications for MWEI as well. Fine-tuned on gold-standard datasets, they show excellent performance, achieving state-of-the-art results (Gombert and Bartsch, 2020; Premasiri and Ranasinghe, 2022). More recently, autoregressive LLMs like GPT or Qwen also started to be used for non-generative tasks, and researchers explored their capability to detect MWEs (Hashiloni et al., 2025; Ide et al., 2025).

However, many endeavors in this field are only tailored to detect MWEs in English (e.g., Schneider et al. (2016)), which has much more training data available in comparison to other languages. Approaches developed as part of PARSEME shared tasks (Savary et al., 2017; Ramisch et al., 2020) or using PARSEME data (Savary et al., 2023) do foster multi-linguality and often consider French MWEs (Bui and Savary, 2024). But PARSEME

data released before 2025 only focused on verbal MWEs, limiting considerably the scope of extraction.

For this reason, the recent publication of the new PARSEME 2.0 corpora,¹ which extend MWE annotation to all grammatical categories, is an important milestone in the MWEI task. Using the French subcorpus of this dataset, we aim to address the existing research gap in the analysis of performance of masked models and LLMs in identifying French MWEs of all types.

3 Data and Models

3.1 Data

We use the updated French subset of the newly released data for PARSEME 2.0 shared task. In comparison to the previous releases, which covered only verbal MWEs, this dataset includes MWEs of all structural types. Since annotated test data was not available at the time this study was carried out, we use the train split for fine-tuning and extraction of few-shot examples and subsequently test the models on the dev split. Table 1 below summarizes some statistics about the data, and Table 6 provided in Appendix B gives an overview of MWE classes represented in the dataset.

Split	Sentences	Tokens	MWEs	MWE Classes
train	3,357	80,559	4,604	16
dev	373	9,353	531	13

Table 1: Data overview.

3.2 Models

We use the large version of CamemBERT² (Martin et al., 2019) as our main model for MWEI task. CamemBERT is based on RoBERTa and has been specifically trained on French data, making it ideal for working with French MWEs. We further fine-tune it for token classification using the train split of the French PARSEME dataset. PARSEME annotations are converted to BIO format, preserving the original tokenization, including split French contractions. Table 3 illustrates how BIO annotation was implemented. Fine-tuning is performed for three epochs using the transformers³ Python

¹https://gitlab.com/parseme/sharedtask-data/-/tree/master/2.0/subtask1?ref_type=heads

²<https://huggingface.co/a1manach/camembert-large>

³<https://pypi.org/project/transformers/>

library, with the best hyperparameters selected via grid search: a learning rate of **5e-05** and a batch size of **16** (see Table 2).

LR	BS	Precision	Recall	F1
2e-05	16	.76	.79	.78
2e-05	32	.69	.73	.71
3e-05	16	.78	.83	.80
3e-05	32	.74	.78	.76
5e-05	16	.79	.82	.81
5e-05	32	.78	.82	.80

Table 2: Grid search results.

Additionally, we use two open-source LLMs for prompt-based MWEI: gpt-oss-20b⁴ (OpenAI, 2025) and Qwen3-32B-AWQ (QwenTeam, 2025),⁵ which we will refer to as gpt-oss and Qwen for brevity. Both models have strong multilingual capabilities, including in French. For instance, Qwen shows similar performance on MMLU (Hendrycks et al., 2021) and MMMLU⁶ benchmarks, which test models’ multitask language understanding in English (MMLU) and other languages (MMMLU). On MMMLU, Qwen outperforms other models of similar size (QwenTeam, 2025). While gpt-oss performs slightly worse on MMMLU in comparison to MMLU, it achieves one of the highest scores in French among languages other than English (OpenAI, 2025).

We prompt both models via the transformers library using the text generation pipeline and setting the temperature to 0.2 for more deterministic responses. Qwen was prompted both with and without reasoning, while for gpt-oss the reasoning was set to medium. The prompt consisted of a system message and a user message. The system message described the models’ role, defined the MWEI task, and provided examples (at least three per MWE type). The user message included the instruction to identify MWEs in the provided target sentence. We tested two versions of the prompt: in the first one, the models were required to return MWEs as pipe-separated strings, while in the second one the models were prompted to annotate each sentence token following the BIO scheme. For the final experiment, we chose the former because the latter resulted in an extremely high proportion of invalid responses. The full final prompt is provided in Appendix A.

⁴<https://huggingface.co/openai/gpt-oss-20b>

⁵<https://huggingface.co/Qwen/Qwen3-32B-AWQ>

⁶<https://huggingface.co/datasets/openai/MMMLU/viewer/default>

Tokens	À	le	delà	de	cette	querelle	,	l'	affaire	est	triste	.
PARSEME	1:AdvID	1	1	*	*	*	*	*	*	*	*	*
BIO	B	I	I	O	O	O	O	O	O	O	O	O

Table 3: Example of BIO labels compared to PARSEME labels. Original tokenization is preserved. English translation of the example: *Beyond this quarrel, the matter is sad.*

4 Results

4.1 MWE Identification

We evaluate the models on MWE level with a conservative procedure (i.e. full match with true MWEs required), considering both macro and category-specific performance. As shown in Table 4, CamemBERT outperforms by large both LLMs, achieving an F1-score of 0.74. Open AI’s gpt-oss ranks higher than Qwen in terms of precision, however achieves considerably lower recall. This may be due to a high proportion of invalid responses delivered by this model. In 128 cases out of 373, it either produced endless repetitions or generated irrelevant text output. For comparison, the proportion of invalid responses by Qwen is notably lower: nine in the run without reasoning and 36 in the run with reasoning. As expected, enabling the reasoning for Qwen improves the performance, albeit only in terms of precision.

Model	Precision	Recall	F1
CamemBERT	.73	.75	.74
gpt-oss	.29	.21	.24
Qwen (no reasoning)	.13	.36	.19
Qwen (with reasoning)	.21	.36	.27

Table 4: Model performance on MWE level (macro).

For the more fine-grained analysis across MWE classes, we grouped subcategories of the same class into a single category. For example, light-verb constructions with semantically bleached verbs (**LVC.full**) and light-verb constructions with causative meaning (**LVC.cause**) were both aggregated under the category of light-verb constructions. Since the task focused on identifying MWEs without assigning specific class labels, we report only recall, computed based on the labels provided in the dataset.

As shown in Figure 1, CamemBERT achieves its highest performance (over 0.8) for adpositional MWEs (**AdpID**), complex conjunctions (**ConjID**), deverbal nominal MWEs (**NV**), and pronominal MWEs (**PronID**). Performance decreases notably for light-verb constructions (**LVC**) and verbal id-

ioms (**VID**), dropping to 0.62 and 0.56, respectively.

Qwen with reasoning enabled performs comparably to CamemBERT in identifying **ConjIDs** and **NVs**. Interestingly, the version with reasoning disabled outperforms the reasoning-enabled variant for all verbal MWEs, as well as for **AdpIDs** and **PronIDs**. In general, Qwen achieves better results than gpt-oss in all but one MWE category.

4.2 Qualitative Error Analysis

In this section, we conduct an error analysis, focusing on false positives, i.e. cases where word sequences identified as MWEs by the models are not considered as such in the dataset. For Qwen, we consider only the results produced with reasoning enabled.

When considering the false positives common to gpt-oss and Qwen, a total of 143 cases is identified. Of these, 70 token sequences are partial matches, usually containing more tokens than those specified in the PARSEME annotations (e.g., *faire confiance à* (to trust + adposition) instead of *faire confiance* (to trust)). In a few cases, both models identified single tokens as MWEs (e.g., *Ajaccio* (Ajaccio), *contenu* (content)), as well as single tokens followed by punctuation (e.g., *comment?* (how?)).

By contrast, CamemBERT false positives, when partially matching PARSEME MWEs, tend to contain fewer tokens than required by the annotation (e.g., *sur la table* (on the table) instead of *être sur la table* (to be on the table)).

The FPs from gpt-oss and Qwen which are not partial matches mainly fall into a small number of recurrent categories (see Table 5). Most of them are noun phrases (NPs), including fixed expressions, domain-specific terminology, institutional names, and proper nouns that do not correspond to PARSEME MWEs. Verb phrases (VPs) and prepositional phrases (PPs) form another important group, typically involving frequent and formulaic constructions. Less frequent false positives include adverbial phrases (AdvPs), conjunctive phrases (ConjPs), and discourse expressions, as well as a small set of miscellaneous cases related to tok-

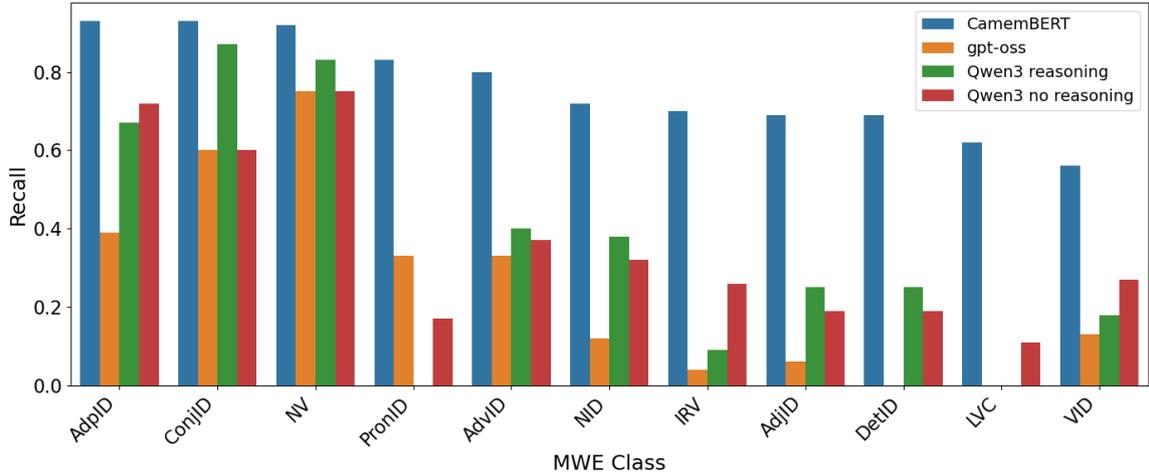


Figure 1: Model recall performance across MWE classes.

FP Class	Examples
NP	<i>prison ferme, durée de conservation, Royaume-Uni, Conseil supérieur de la magistrature</i>
VP	<i>faire aboutir, se focaliser sur, nous contenter de, se concentrer sur</i>
PP	<i>sur le fondement de, sur le thème de, non loin de, en faveur de</i>
AdvP	<i>une nouvelle fois, dans son ensemble, fort heureusement, mercredi soir</i>
ConjP	<i>c'est pourquoi, ni l'un ni l'autre</i>
DiscE	<i>voir aussi, chers collègues, monsieur le président</i>
Other	<i>de le, eu/1/04/289/001, comment ?</i>

Table 5: Main categories of false positives produced by GPT and Qwen, with representative examples for each class.

enization or metadata.

When restricting the analysis to CamemBERT false positives that are not partial matches, most errors correspond to isolated lexical units or complete but non-idiomatic phrases. A large share consists of noun phrases, including abstract nouns, technical terms, and institutional or domain-specific expressions (e.g., *secteur économique* (economic sector), *aide sociale* (social assistance), *gouvernement fantôme* (shadow government)). Another common type consists of prepositional or adverbial phrases that are syntactically well-formed collocations but are not considered MWEs under the PARSEME framework (e.g., *en fuite* (on the run), *dans les délais* (on time / within the deadline), *non loin de* (not far from)). CamemBERT also predicts single verbs or function words and truncated elements that do not constitute MWEs (e.g., *laisser* (to let), *parlez* (speak / talk [imperative or plural]), *il* (he/it), *de* (of/from), *s'* (oneself / reflexive clitic)).

5 Discussion

Our findings confirm the initial expectation that fine-tuned encoder-only language models outperform pretrained LLMs on the MWEI task. No-

tably, even fine-tuning LLMs or employing complex prompting setups, such as chain-of-thought reasoning, are unlikely to achieve performance comparable to that of encoder-only models (see, for example, the scores reported by Hashiloni et al. (2025) on the CoAM dataset, which is similar to PARSEME corpus). Besides, it is worth noting that our fine-tuning approach was rather straightforward; more sophisticated strategies could potentially result in even higher performance (cf. Bui and Savary (2024)).

As shown in the previous section, all models are better in identifying **AdpID**, **ConjID**, and **NVs**. The first two of these classes are arguably the most highly grammaticalized MWEs, which likely facilitates recognition by models. In contrast, the lowest performance is observed for verbal MWEs (**VID** and **LVC**), which show greater variability and are frequently realized in discontinuous forms, posing a substantial challenge for MWEI systems (cf. Bui and Savary (2024); Constant et al. (2017)).

The error analysis of false positives reveals that, due to the large number of partial matches, one of the main challenges lies in correctly defining MWE boundaries. Moreover, the presence of discontinu-

ous forms may also explain the substantial number of false positives consisting of single tokens.

6 Conclusion and Future Work

In this study, we compared the performance of fine-tuned CamemBERT with two LLMs (gpt-oss-20b and Qwen3-32B-AWQ) in the task of automatic identification of French MWEs. Our results show that the fine-tuned model outperforms both LLMs by a large margin, although Qwen shows comparable performance in the identification of some MWE types. Our error analysis reveals that CamemBERT tends to underpredict MWE tokens, while both LLMs generally return longer MWE sequences, often including function words in MWEs. Future work will incorporate multi-lingual masked models and a wider selection of LLMs, including LLMs predominantly trained on French data. Additionally, we will explore if using more fine-grained label schemes can improve model performance.

Limitations

This study is subject to some limitations that should be taken into account. First, our evaluations were based on the dev data split due to the unavailability of annotated test data. Second, this study used a conservative evaluation strategy requiring a full match between a true MWE and model predictions, which penalises model performance. Nevertheless, while a token-based evaluation will yield higher scores, the margin is likely to remain in favour of CamemBERT. Third, LLMs were prompted only in a few-shot setting. Testing other prompting strategies would give a more complete understanding of LLM capabilities for the identification of French MWEs.

Acknowledgments

This research is funded by *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102.

References

Kamyar Arzideh, Henning Schäfer, Héctor Allende-Cid, Giulia Baldini, Thomas Hilser, Ahmad Idrissi-Yaghir, Katharina Laue, Nilesh Chakraborty, Niclas Doll, Dario Antweiler, Katrin Klug, and Niklas Beck et al. 2025. [From BERT to Generative AI - Comparing Encoder-only vs. Large Language Models in a Cohort of Lung Cancer Patients for Named Entity Recognition in Unstructured Medical Reports](#). *Computers in Biology and Medicine*, 195:110665.

Abayomi Bello, Sin-Chun Ng, and Man-Fai Leung. 2023. [A BERT Framework to Sentiment Analysis of Tweets](#). *Sensors*, 23(1).

Martin Bucher and Marco Martini. 2024. [Fine-Tuned ‘Small’ LLMs \(Still\) Significantly Outperform Zero-Shot Generative AI Models in Text Classification](#). *ArXiv*, abs/2406.08660.

Van-Tuan Bui and Agata Savary. 2024. [Cross-type French Multiword Expression Identification with Pre-trained Masked Language Models](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4198–4204, Torino, Italia. ELRA and ICCL.

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

Eduardo C. Garrido-Merchan, Roberto Gozalo-Brizuela, and Santiago Gonzalez-Carvajal. 2023. [Comparing BERT Against Traditional Machine Learning Models in Text Classification](#). *Journal of Computational and Cognitive Engineering*, 2(4):352–356.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. [ChatGPT Outperforms Crowd Workers for Text-annotation Tasks](#). *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.

Sebastian Gombert and Sabine Bartsch. 2020. [MultiVitaminBooster at PARSEME Shared Task 2020: Combining window- and dependency-based features with multilingual contextualised word embeddings for VMWE detection](#). In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, pages 149–155, online. Association for Computational Linguistics.

Kai Golan Hashiloni, Ofri Hefetz, and Kfir Bar. 2025. [Easy as PIE? Identifying Multi-Word Expressions with LLMs](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 23782–23801, Suzhou, China. Association for Computational Linguistics.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. [Measuring massive multitask language understanding](#). *Preprint*, arXiv:2009.03300.

Yusuke Ide, Joshua Tanner, Adam Nohejl, Jacob Hoffman, Justin Vasselli, Hidetaka Kamigaito, and Taro Watanabe. 2025. [CoAM: Corpus of All-Type Multiword Expressions](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 27004–27021, Vienna, Austria. Association for Computational Linguistics.

Kai Labusch, Preußischer Kulturbesitz, Clemens Neudecker, and David Zellhöfer. 2019. BERT for

- Named Entity Recognition in Contemporary and Historical German. In *Proceedings of the 15th conference on natural language processing*, pages 9–11.
- Louis Martin, Benjamin Muller, Pedro Ortiz Suarez, Yoann Dupont, Laurent Romary, Eric Villemonte de la Clergerie, Djamé Seddah, and Benoît Sagot. 2019. [CamemBERT: a Tasty French Language Model](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Juan G. Diaz Ochoa, Natalie Layer, Jonas Mahr, Faizan E. Mustafa, Christian U. Menzel, Martina Müller, Tobias Schilling, Gerald Illerhaus, Markus Knott, and Alexander Krohn. 2025. [Optimized BERT-based NLP Outperforms Zero-shot Methods for Automated Symptom Detection in Clinical Practice](#). *Frontiers in Digital Health*, 7:1623922.
- OpenAI. 2025. [gpt-oss-120b & gpt-oss-20b Model Card](#). *Preprint*, arXiv:2508.10925.
- Damith Premasiri and Tharindu Ranasinghe. 2022. [BERT\(s\) to Detect Multiword Expressions](#). *Preprint*, arXiv:2208.07832.
- QwenTeam. 2025. [Qwen3 Technical Report](#). *Preprint*, arXiv:2505.09388.
- Carlos Ramisch, Agata Savary, Bruno Guillaume, Jakub Waszczuk, Marie Candito, Ashwini Vaidya, Verginica Barbu Mititelu, Archana Bhatia, Uxoá Inurrieta, Voula Giouli, and Tunga et al. Güngör. 2020. [Edition 1.2 of the PARSEME Shared Task on Semi-supervised Identification of Verbal Multiword Expressions](#). In *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, pages 107–118, online. Association for Computational Linguistics.
- Agata Savary, Cherifa Ben Khelil, Carlos Ramisch, Voula Giouli, Verginica Barbu Mititelu, Najet Hadj Mohamed, Cvetana Krstev, Chaya Liebeskind, Hongzhi Xu, Sara Stymne, Tunga Güngör, Thomas Pickard, Bruno Guillaume, Eduard Bejček, Archana Bhatia, Marie Candito, and Polona et al. Gantar. 2023. [PARSEME corpus release 1.3](#). In *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*, pages 24–35, Dubrovnik, Croatia. 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.
- Nathan Schneider, Dirk Hovy, Anders Johannsen, and Marine Carpuat. 2016. [SemEval-2016 Task 10: Detecting Minimal Semantic Units and their Meanings \(DiMSUM\)](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 546–559, San Diego, California. Association for Computational Linguistics.
- Miklós Sebők, Viktor Kovács, Martin Bánóczy, Daniel Møller Eriksen, Nathalie Neptune, and Philippe Roussille. 2025. [Beyond token limits: Assessing language model performance on long text classification](#). *Preprint*, arXiv:2509.10199.
- Anna Siyanova-Chanturia. 2013. Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, 8(2):245–268.
- Anna Siyanova-Chanturia, Kathy Conklin, Sendy Cafarra, Edith Kaan, and Walter JB van Heuven. 2017. Representation and processing of multi-word expressions in the brain. *Brain and language*, 175:111–122.
- Petter Törnberg. 2023. [ChatGPT-4 Outperforms Experts and Crowd Workers in Annotating Political Twitter Messages with Zero-Shot Learning](#). *Preprint*, arXiv:2304.06588.

A LLM Prompt

System message:

#Role

You are a professional annotator with strong linguistic background. Your area of specialization is multi-word expressions (MWEs).

#Task

MWEs are sequences of at least two words that show some degree of orthographic, morphological, syntactic and/or semantic idiosyncrasy with respect to what is considered general grammar rules of a language. Your task will be to identify MWEs in French texts.

#Input

You will be provided with a sentence in French that may or may not contain MWEs.

#Instructions

Extract MWEs detected in the sentence if any. If more than one MWE is detected, separate the MWEs with a pipe (|). MWEs can be discontinuous, i.e. contain optional tokens that are not part of an MWE. Do not extract such tokens.

#Output

Return the extracted MWEs. If no MWEs detected, return the original sentence. Do not include any additional text or comments.

#Examples

FEW-SHOT EXAMPLES

User message:

Identify MWEs in the following sentence: TARGET SENTENCE

B MWE Classes

MWE Class	Description	Examples
LVC.cause	Causative light-verb constructions	<i>susciter intérêt, entraîner réduction, apporter cohérence</i>
LVC.full	Light-verb constructions with bleached verbs	<i>avoir droit, faire appel, donner instructions</i>
VID	Verbal idioms	<i>venir en aide, donner un blanc-seing, avoir lieu</i>
IRV	Inherently reflexive verbs	<i>s'apparenter, se trouver, se rendre</i>
MVC	Multi-verb constructions	<i>faire valoir, laisser faire, entendre parler</i>
NID	Nominal idioms	<i>homme d'affaires, chaîne de radio, droit de vote</i>
NV.VID	Deverbal nominal MWEs stemming from a VID	<i>mise au point, retour à la normale, entrée en vigueur</i>
NV.MVC	Deverbal nominal MWEs stemming from an MVC	<i>savoir-faire</i>
NV.LVC.full	Deverbal nominal MWEs stemming from an LVC.full	<i>ayant droit, prises de position</i>
AdjID	Adjectival idioms	<i>à part entière, en désaccord, de premier plan</i>
AdvID	Adverbial idioms	<i>en même temps, en principe, c'est-à-dire</i>
AdpID	Adposition idioms	<i>grâce à, lors de, aux termes de</i>
ConjID	Conjunction idioms	<i>parce que, à moins que, ou bien</i>
DetID	Determiner idioms	<i>plein de, majorité de, un peu de</i>
IntjID	Interjection idioms	<i>eh bien, bonne chance</i>
PronID	Pronominal idioms	<i>elle-même, tout le monde, quelque chose</i>

Table 6: MWE classes in French PARSEME data.