LLM as a Guide: an Approach for Unsupervised Economic Relation Discovery in Administrative Documents

Thomas Sebbag[1,2], Solen Quiniou[1], Emmanuel Morin[1],

[1] Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France,
[2] Explore, Carquefou, France,
Correspondence: thomas.sebbag@univ-nantes.fr, solen.quiniou@univ-nantes.fr, emmanuel.morin@univ-nantes.fr

Abstract

Effective relation extraction (RE) from unstructured text is critical, especially when the target relations are unknown. In this case, we can leverage Large Language Models (LLMs) to perform this task. In this paper, we present an LLM as a guide approach for identifying economic relations. This novel methodology, based on sentence clustering and an LLM, is used to identify previously unknown economic relations in French administrative documents. It addresses the challenge of extracting actionable knowledge without predefined relation labels. We evaluate our approach on French and English RE datasets, demonstrating high precision and recall in detecting previously unknown relations. Our results suggest that clustering and LLM-based methods can effectively discover and categorize economic relations, with potential applications to private corpora.

1 Introduction

Effective information extraction poses a significant challenge when the target information is undefined. Identifying the relevance of the information to be extracted is a critical step, particularly in the case of analyzing market dynamics and organizational decisions in the economic domain. As a subfield of the economic domain, the administrative domain is interested in the textual productions of various public organizations in order to analyze the behavior of economic actors from the administration's perspective.

On the one hand, Open Relation Extraction (ORE) aims to transform unstructured text into structured and actionable knowledge by using an unsupervised mechanism to extract predefined relations as presented by Shukla et al. (2025). According to Jiang et al. (2024), traditional RE and ORE methods mainly focus on predefined patterns that refer to a predefined set of relations and entities given during the training step. ORE primarily rely on two datasets from traditional RE: TACRED (Zhang et al., 2017) and Fewrel (Han et al., 2018). These datasets were created from English data from news wires and web text, and the relations present in these datasets are clearly defined. However, to the best of our knowledge, few studies address the problem of identifying unknown relations, and even fewer propose a methodology for exploring a corpus specifically dedicated to this task. Indeed, extracting relations is even more challenging when the target information to be extracted is undefined.

On the other hand, Semantic Typing (ST) attempts to assign tokens or relevant text spans to semantic categories such as relation types, entity types, or event types within a given context (Huang et al., 2022). This task may correspond to the first stage of the RE process. We aim to apply this approach to the administrative domain by constructing a relational schema of the interactions between public administrations and their environment, based on their textual production, to support economic analysis.

In this work, we propose a methodology inspired by Wrzalik et al. (2024) to identify unknown economic relations, in French documents produced by a public administration. Our contributions can be summarized as follows:

- An approach based on sentence clustering and an LLM guiding experts in the process of identifying unknown relations;
- An evaluation of our methodology on two RE datasets (one in French and one in English) to assess its ability to capture general relations.
- The data and associated relations from our administrative corpus[1].

[1] Github repository

Proceedings of The 10th Workshop on Financial Technology and Natural Language. EMNLP-2025, Suzhou, China

2 Related Work

2.1 Open Relation Extraction

Information Extraction (IE) extracts structured relationships in the format < *arg1; rel; arg2* > using semantic or syntactic cues to classify sequences (Niklaus et al., 2018). ORE extends this definition to identify new relations from unlabeled open-domain corpora (Wang et al., 2024). While these methods provide a basis, they struggle to capture global context.

Clustering-based approaches (Wang et al., 2022; Li et al., 2022; Zhou et al., 2023) are another method for addressing RE by grouping relations into types through Masked Language Modeling or feature extraction. However, these are sensitive to data quality, dataset size, and relation distribution, making them reliant on the training data, especially for sharp topics.

Generative Relation Extraction (GRE) is an LLM-based approach that comprehends input text and identifies relations in a zero-shot setting, without relying on predefined patterns (Jiang et al., 2024). LLMs in zero-shot and few-shot setups have been shown to perform GRE without fine-tuning (Wadhwa et al., 2023; Li et al., 2023) and have been used to extract economic relations (Ettaleb et al., 2025). However, GRE methods still rely on a predefined set of relations and entities, somewhat similar to traditional RE. Wrzalik et al. (2024) uses a methodology based on statement retrieval and LLM to extract text passages about companies emmisions goals. Inspired by this approach, we propose an approach to discovering unknown relations in a zero-shot setup without predefined relation definition.

2.2 Semantic Typing

Although relation classification aims at categorizing defined relations, it remains a close task to semantic typing, a task focusing on identifying unknown relations according to Thomas et al. (2024). Recent works rely on the typing technique to discover unknown relations, using entity types as a preliminary stage to relation identification on TACRED and French press documents (Lyu and Chen, 2021; Mallart et al., 2021).

3 Datasets

To identify unknown relations between economic actors in a given area and conduct an economic analysis, we use French administrative documents. Potin et al. (2023) and Sebbag et al. (2025) described the textual production of public administrations as a great playground to harvest economic information. They also described these documents as challenging for NLP tasks due to their lack of structure and the potential amount of noise. In this section, we present the three datasets considered.

3.1 Administrative Data

In public administrations, decision documents summarizing actions and debates by authorities are particularly valuable. Discussions with domain experts confirmed that those documents contain key economic relations, offering insights for market understanding.

Following expert recommendations, we selected the French city of Lambesc to initiate our relation typing study. To ensure consistency, we focused on decisions from 2024, a year with an unchanged administrative team. The resulting corpus includes 830 unique sentences and 2,460 named entities (*ORG*, *LOC*, *PER*, in that order; see Appendix A.1). We used internal tools to scrape publicly available documents from Lambesc's website, extracting only sentences to simplify this initial exploratory analysis.

Example of extracted sentence translated to English : *The Town Hall of Lambesc informs the assembly that the commune has applied to the SAFER for the acquisition of the plot of land cadastral section AT n°84 located in Bonrecueil Nord.*

3.2 Other Relation Extraction Datasets

To evaluate our methodology, we also chose two annotated datasets: one close to our domain, containing French economic relations (BizRel[2]), and one containing general English relations (FewRel[3]).

Bizrel (Khaldi et al., 2022) is a multilingual dataset focusing on Business Relation extraction between organizations. It contains 2 007 sentences in French and six relations including 5 economic relations (details in Appendix A.2). Only the French part of the dataset was used in our experiment. To the best of our knowledge, Bizrel is the open source dataset that most closely matches our use case in the context of extracting economic relations.

FewRel (Han et al., 2018), is an English Relation Extraction dataset consisting of 100 general

[2]https://github.com/Geotrend-research/business-relation-dataset

[3]https://github.com/thunlp/FewRel

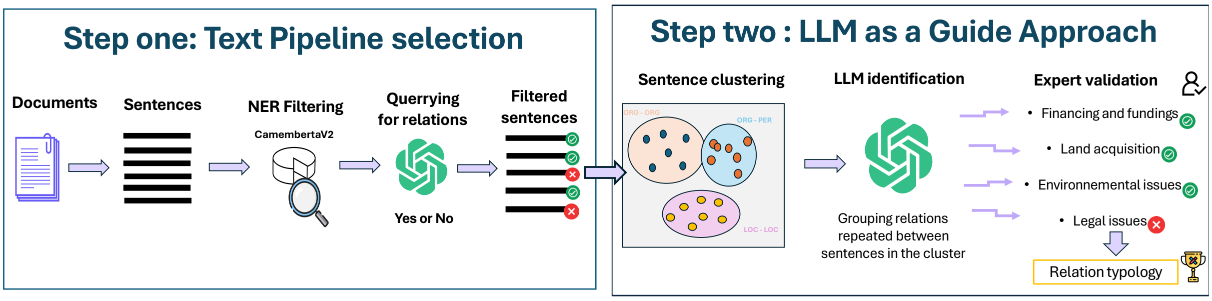

Figure 1: Our two-step pipeline for relation identification, with text selection and LLM-based relation generation.

domain relations from Wikipedia data. We created a random snapshot of 1 611 sentences, from the validation set. This snapshot contains 16 relations described in Appendix A.3. The entity types in this dataset match ours, making it an appropriate resource for evaluating our setup on English data.

4 Relation Exploration Pipeline

In this section, we describe the two steps of our relation identification pipeline, on administrative documents, as shown in Figure 1.

4.1 Step One: Text Pipeline Selection

To identify unknown relations, we go through a phase of selecting sentences that might contain a relation.

The first sub-step consists of segmenting the texts into sentences, based on the hypothesis that sentence-level relations may be more explicit, making automatic relation identification easier. For this text segmentation, we use Langchain[4] module *tiktoken*, as well as Spacy Sentencizer[5] with the *fr_core_news_sm* model.

In a second sub-step, we detect named entities using a CamemBERTaV2 language model (Antoun et al., 2024), fine-tuned on Adminset-NER (Sebbag et al., 2025), an annotated NER dataset consisting exclusively of French administrative documents. This allows us to filter sentences to retain only those that contain at least two named entities

The last sub-step is to use an LLM for a binary task of detecting whether or no a relation is present within the sentence. GPT-4o (OpenAI et al., 2024) was chosen for this task because of its ability to highlight potential relations, as suggested by Ding et al. (2024).

[4]https://python.langchain.com
[5]https://spacy.io/api/sentencizer

4.2 Step Two: LLM-as-a-Guide Approach

Our method proposes an approach based on using an LLM as a guide to identify unknown relations in sentences, after regrouping them into clusters.

For the first sub-step, we created clusters using entity pair templates which could contain a relation.For example, a single cluster will contain all sentences that have both an ORG entity and a LOC entity. We also evaluate this clustering method to a random selection of sentences.

In the next sub-step, we prompted the LLM with a temperature of 0.001 to identify the relation types that are repeated in the sentences of each given cluster, providing the associated texts and entities. (See Appendices A.4 and A.5 for the prompts in French and English).

In the final step, three experts collaboratively evaluated the relations proposed by the LLMs. They serve as Product Owners or Product Managers at a company specialized in information delivery for public procurement applications. Reaching consensus on whether each relation was valid, invalid, or overlapping with a previously identified one was required for validation.

5 Experiments and Results

In this section, we present the experimental setup and the results on the different datasets considered.

Table 1 and 2 present the results, in terms of precision (P) recall (R or R') and f1-score (F1) on relation types.

5.1 Relation Identification on Administrative Data

For our first experiment, we applied entity pair clustering to administrative data (Admin Entity), generating five clusters based on predefined templates. We compared this to random sentence selection (Admin Random), which formed four clusters while maintaining a balanced ratio of repeated

Models	Admin Random		Admin Entity	
	P	R'	P	R'
GPT-4o	88,88%	57.14%	85,71%	85.71%
GPT-4.1	**91,67%**	**78.57%**	**87,50%**	**100.00%**
Llama 3.1 70B	80,00%	28.57%	55,56%	71.43%
Mixtral 8*22B	66,67%	28.57%	60,00%	64.29%
Llama 3.1 8B	40.00%	28.57%	36,84%	50,00%
Mixtral 7B	42.86%	21,43%	54,55%	42,86%

Table 1: Precision and an alternative version of Recall results for relation identification on administrative data, on two methods used to create sentence clusters.

relations. Since relations between entities are not known in advance, precision (P) is calculated based on those judged relevant by domain experts. To address the lack of annotated ground truth, we introduce an alternative recall metric (R'), based on the maximum number of relations, 14 in total, generated by GPT-4.1 and validated by domain experts.

Admin Random As shown in Table 1, GPT models exhibits strong zero-shot performance, with GPT-4.1 achieving 91.66% precision but a recall of 78,57% identifying 11 relations. Only one false positive was noted, where a legal reminder was misclassified as a relation. Llama 3.1 70B and Mixtral 8x22B each produced only four relations, mostly generic, less than half of GPT-4.1's output, with two false positives for Mixtral 8x22B. Llama 3.1 8B also produced four relations, but created seven false positives, most of which corresponded to inaccurate labels regarding the provided examples. Mistral 7B obtained the lowest results with only generic labels. Only three of the labels were identified correctly, while four were identified as false positives, which explains the low recall.

Admin Entity In this setup, GPT models keep their advance on relation identification. While precision exhibits a slight decrease with a score of 87.45%, recall presents a perfect score, generating the most true relations on topics like *project financing*, *land transfer/acquisition*, and *city planning* (see Appendix A.7), with only 2 identified as false positives. Llama 3.1 70B and Mixtral 8x22B kept struggling in this setup, generating general relations with a high proportion of false positives; however, the models improved the quality of the labels generated, even though some of them were too generic. The smaller models improved their average performance in terms of recall, but Llama 3.1 8B over generated twelve false labels while Mistral 7B generated a significant number of duplicate relations.According to domain experts, the relations generated in this setup were generally finer than before and encapsulated more precise semantic concepts by using entity pairs to create clusters.

In conclusion, experts most appreciated GPT's ability to identify a wide range of relations and generate clear, meaningful labels, which reflect in our version of the recall score.

5.2 Relation Identification on Annotated Datasets

Datasets	Models	P	R	F1
Bizrel	GPT-4o	83,33%	83,33%	**83,33%**
	GPT-4.1	83,33%	83,33%	**83,33%**
	Llama 3.1 70B	66,67%	66,67%	66,67%
	Mixtral 8*22B	50,00%	66,67%	57,14%
	Llama 3.1 8B	100,00%	66,67%	80,00%
	Mistral 7B	60,00%	50,00%	54,55%
Fewrel	GPT-4o	82,35%	87,50%	**84,85%**
	GPT-4.1	81,25%	81,25%	81.25%
	Llama 3.1 70B	57,14%	25,00%	34,78%
	Mixtral 8*22B	64,29%	52,94%	58,06%
	Llama 3.1 8B	26,32%	31,25%	28,57%
	Mistral 7B	50,00%	18,75%	27,27%

Table 2: Results for relation identification on annotated datasets.

As shown in Table 2, experiments on BizRel and FewRel used random sentence selection to form four clusters. Entity pair clustering was not applicable, as BizRel contains only *ORG* entities and FewRel does not provide entity labels in its validation set. Since both datasets contain annotated relations, we evaluated our protocol using standard metrics precision (P), recall (R), and F1-score (F1).

Results on BizRel This dataset includes six relation types, one of which is *Other*. GPT models achieved an F1 score of 83.33%, correctly identifying all five economic relations. However, they introduced a non-existent type, *Ranking comparison*, and failed to capture the *Other* category. In contrast, Llama 3.1 70B and Mixtral 8x22B struggled to identify accurate relations, often misinterpreting the context. Interestingly, the smaller models outperformed their larger counterparts in terms of precision during this experiment. Llama 3.1 8B benefited from this advantage, achieving a higher F1 score by producing only one false positive. Conversely, Mistral 7B underperformed in terms of F1 score, generating three positive relations and two false positives.

Results on FewRel The FewRel validation snapshot includes 16 relation types. GPT models again achieved the best performance, with GPT-4o reaching an F1 score of 84.85%, correctly identifying 13 out of 16 relations. However, GPT's models consistently struggled with certain types, such as *mother*, *main subject*, and *voice type*. Llama 3.1 70B predicted only seven relations, four of which were correct, resulting in a low recall. Mixtral 8x22B demonstrated a stronger ability to identify relevant relations but frequently produced incorrect labels, reflecting difficulties in capturing contextual nuances. Notably, its outputs included unusually accurate label generation, which could indicate potential data contamination given the model's overall performance across our experiments. However, this remains a hypothesis requiring further investigation. Of the smaller models, Llama 3.1 8B generated five correct relations but produced fourteen false positives, suggesting that it was overwhelmed by the number of relations to identify. Mistral 7B obtained the lowest F1 score, producing only three true positives and duplicating labels across clusters.

GPT models demonstrated strong overall performance across both evaluation datasets but also exhibited consistent weaknesses in recognizing specific relation types. These results suggest potential biases toward certain subjects or difficulty detecting relations that differ greatly from the training data.

6 Conclusion

In this work, we propose an approach based on sentence clustering and an LLM guiding experts in the process of identifying unknown relations in the French administrative domain. We also evaluated our methodology on two annotated datasets for relation extraction in French and English. The validation by industry experts highlights its potential for economic analysis. Furthermore, this method could be applied to private company corpora, as they share similar unstructured frameworks and domain-specific terms. Our experiments mark a first step towards extracting complex relations from administrative documents, with plans to extend it to complex paragraphs and implicit information. Ultimately, we aim to create an annotated corpus for relation extraction, and we hope that our methods will inspire future work in relation identification and information retrieval.

Limitations

It seems important to us to discuss some of the limitations of our experiments:

- Given that FewRel is an open domain dataset obtained from the web, we are concerned about potential data contamination during LLMs training process, which could affect its performance in our experiment. Although we did not evaluate the model directly on the RE task, but rather on its ability to generate relation labels, it is still possible to use its preexisting knowledge base, including this data, to produce more accurate results.

- To effectively evaluate LLMs on our administrative data, we chose to calculate recall based on the maximum number of correct relations generated by GPT-4.1, given that the data are not yet annotated at this stage of the project. This approach introduces a bias in favor of GPT models, which can hinder independent verification. Three domain experts evaluated the relations, and we consider their expertise sufficient to assess the relevance of each relation and base an evaluation score on it.

- Our approach relies on proprietary models, raising concerns about dependency on private companies. Furthermore, these models only partially reveal their inner workings, which restricts our ability to analyze their outputs.

Ethical Concerns

All datasets used in this study, BizRel and FewRel, are publicly available. The French administrative data from the Lambesc website are available under a Creative Commons license under the Attribution-NonCommercial 4.0 International License. This transparency minimizes ethical concerns regarding data acquisition and usage.

In addition, the interpretability and transparency of LLM's decision-making processes are essential. Recognizing the limitations and biases of LLMs, including occasional information inaccuracies, we emphasize the importance of reliability in our evaluation methodology. Furthermore, the integration of LLMs-as-a-guide impacts traditional human roles, requiring careful consideration of the ethical implications of labor displacement. Moreover, the powerful capabilities of LLMs underscore the need for responsible use and measures to prevent misuse, aligning our research with high ethical standards and societal well-being. We carefully reviewed and ensured that the data we used to input each LLM did not contain any offensive information.

The total cost of our experiments is estimated to be 48.86$, according to OpenAI's billing history regarding the usage of GPT models. The total environmental cost, according to the Jean Zay supercomputer documentation is equivalent to 58.872 Wh or 3.07 kg CO2eq based on the carbon intensity of the energy grid mention by BLOOM environmental cost study also made on Jean Zay (Luccioni et al., 2022).

References

Wissam Antoun, Francis Kulumba, Rian Touchent, Éric de la Clergerie, Benoît Sagot, and Djamé Seddah. 2024. Camembert 2.0: A smarter french language model aged to perfection. *Preprint*, arXiv:2411.08868.

Linyi Ding, Jinfeng Xiao, Sizhe Zhou, Chaoqi Yang, and Jiawei Han. 2024. Topic-Oriented Open Relation Extraction with A Priori Seed Generation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13834–13845, Miami, Florida, USA. Association for Computational Linguistics.

Mohamed Ettaleb, Véronique Moriceau, Mouna Kamel, and Nathalie Aussenac-Gilles. 2025. The contribution of LLMs to relation extraction in the economic field. In *Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal) - Proceedings of the workshops*, pages 175–183, Abu Dhabi, United Arab Emirates. Association of COmputational Linguistics. Backup Publisher: Chung-Chi Chen and Antonio Moreno-Sandoval and Qianqian Xie and Jimin Huang and Sophia Ananiadou and Hsin-Hsi Chen.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4803–4809, Brussels, Belgium. Association for Computational Linguistics.

James Y. Huang, Bangzheng Li, Jiashu Xu, and Muhao Chen. 2022. Unified Semantic Typing with Meaningful Label Inference. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2642–2654, Seattle, United States. Association for Computational Linguistics.

Pengcheng Jiang, Jiacheng Lin, Zifeng Wang, Jimeng Sun, and Jiawei Han. 2024. GenRES: Rethinking evaluation for generative relation extraction in the era of large language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2820–2837, Mexico City, Mexico. Association for Computational Linguistics.

Hadjer Khaldi, Farah Benamara, Grégoire Siegel, Camille Pradel, and Nathalie Aussenac-Gilles. 2022. How's Business Going Worldwide ? A Multilingual Annotated Corpus for Business Relation Extraction. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pages 3696–3705, Marseille, France. European Language Resources Association (ELRA).

Guozheng Li, Peng Wang, and Wenjun Ke. 2023. Revisiting large language models as zero-shot relation extractors. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6877–6892, Singapore. Association for Computational Linguistics.

Sha Li, Heng Ji, and Jiawei Han. 2022. Open relation and event type discovery with type abstraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6864–6877, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Alexandra Sasha Luccioni, Sylvain Viguier, and Anne-Laure Ligozat. 2022. Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model. *arXiv preprint*. ArXiv:2211.02001 [cs].

Shengfei Lyu and Huanhuan Chen. 2021. Relation Classification with Entity Type Restriction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 390–395, Online. Association for Computational Linguistics.

Cyrielle Mallart, Michel Le Nouy, Guillaume Gravier, and Pascale Sébillot. 2021. Active learning for interactive relation extraction in a French newspaper's articles. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 886–894, Held Online. INCOMA Ltd.

Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2018. A survey on open information extraction. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3866–3878, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and

262 others. 2024. GPT-4 Technical Report. *arXiv preprint*. ArXiv:2303.08774 [cs].

Lucas Potin, Vincent Labatut, Pierre-Henri Morand, and Christine Largeron. 2023. FOPPA: an open database of French public procurement award notices from 2010–2020. *Scientific Data*, 10(1):303. Publisher: Nature Publishing Group.

Thomas Sebbag, Solen Quiniou, Nicolas Stucky, and Emmanuel Morin. 2025. AdminSet and AdminBERT: a Dataset and a Pre-trained Language Model to Explore the Unstructured Maze of French Administrative Documents. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 392–406, Abu Dhabi, UAE. Association for Computational Linguistics.

Neelesh Kumar Shukla, Prabhat Prabhakar, Sakthivel Thangaraj, Sandeep Singh, Weiyi Sun, C Prasanna Venkatesan, and Viji Krishnamurthy. 2025. GraphRAG Analysis for Financial Narrative Summarization and A Framework for Optimizing Domain Adaptation. In *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*, pages 23–34, Abu Dhabi, UAE. Association for Computational Linguistics.

Hugo Thomas, Guillaume Gravier, and Pascale Sébillot. 2024. Recherche de relation à partir d'un seul exemple fondée sur un modèle N-way K-shot : une histoire de distracteurs. In *Actes de la 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 1 : articles longs et prises de position*, pages 157–168, Toulouse, France. ATALA and AFPC.

Somin Wadhwa, Silvio Amir, and Byron Wallace. 2023. Revisiting relation extraction in the era of large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15566–15589, Toronto, Canada. Association for Computational Linguistics.

Jiaxin Wang, Lingling Zhang, Wee Sun Lee, Yujie Zhong, Liwei Kang, and Jun Liu. 2024. When phrases meet probabilities: Enabling open relation extraction with cooperating large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13130–13147, Bangkok, Thailand. Association for Computational Linguistics.

Jiaxin Wang, Lingling Zhang, Jun Liu, Xi Liang, Yujie Zhong, and Yaqiang Wu. 2022. MatchPrompt: Prompt-based open relation extraction with semantic consistency guided clustering. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7875–7888, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Marco Wrzalik, Florian Faust, Simon Sieber, and Adrian Ulges. 2024. NetZeroFacts: Two-Stage Emission Information Extraction from Company Reports. In *Proceedings of the Joint Workshop of the 7th Financial Technology and Natural Language Processing, the 5th Knowledge Discovery from Unstructured Data in Financial Services, and the 4th Workshop on Economics and Natural Language Processing*, pages 70–84, Torino, Italia. Association for Computational Linguistics.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware Attention and Supervised Data Improve Slot Filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.

Sizhe Zhou, Suyu Ge, Jiaming Shen, and Jiawei Han. 2023. Corpus-Based Relation Extraction by Identifying and Refining Relation Patterns. In *Machine Learning and Knowledge Discovery in Databases: Research Track*, pages 20–38, Cham. Springer Nature Switzerland.

A Appendix

A.1 Detailed statistics on our administrative corpus

	Sentences	Entities		
		ORG	LOC	PER
Admin Data	830	1 325	738	397

Table 3: Statistics of the administrative dataset.

A.2 BizRel relations

Labels	Details
Investment	An organisation is a subsidiary of another organisation, or an organisation holds (all or part) of the shares of another organisation.
Competition	A competition/rivalry between two organisations providing the same goods or services, or wanting to access the same relatively small market.
Cooperation	A contractual cooperation between two organisations, or when two organisation work together on the same project.
Legal Proceedings	One organisation launches a legal proceedings against another organisation.
Sale-Purchase	One organisation is a client of another, or supplies it with goods or services.
Others	If none of the previously described relations are expressed between the tagged entity pair, or if other types of relations out of this list are expressed, the relation should be OTHERS.

Table 4: Details on the relations in the BizRel dataset.

A.3 FewRel relations

Labels	Details
Crosses	Obstacle (body of water, road, etc) which this bridge crosses over or this tunnel goes under
Original language of film or TV show	Language in which a film or a performance work was originally created. Or language of work or name.
Competition class	Official classification by a regulating body under which the subject (events, teams, participants, or equipment) qualifies for inclusion.
Part of	Object of which the subject is a part.
Sport	Sport in which the subject participates or belongs to.
Constellation	The area of the celestial sphere of which the subject is a part.
Position played on team/speciality	Position or specialism of a player on a team.
Located in or next to body of water	Body of water on or next to which a place is located.
Voice type	Person's voice type. expected values: soprano, mezzo-soprano, contralto, countertenor, tenor, baritone, bass (and derivatives).
Follows	Immediately prior item in a series of which the subject is a part.
Spouse	The subject has the object as their spouse (husband, wife, partner, etc.).
Military rank	Military rank achieved by a person, or military rank associated with a position.
Mother	Female parent of the subject.
Member of	Organization or club to which the subject belongs.
Child	Subject has object as biological, foster, and/or adoptive child.
Main subject	Primary topic of a work.

Table 5: Details on the relations in the FewRel dataset.

A.4 [French version] Prompt used for explore relations into administrative documents with LLMs

```
Tu es un assistant IA qui est chargé de faire de l'extraction de relations
dans des batch de textes écrient par des administrations publique française.

Tu reçois en entré plusieurs textes ayant été regroupé dans la même catégorie
ainsi que les entités nommées présentent dans ces textes entre crochet,
tu dois déterminer si ces textes sont reliés par des relations similaires
en indiquant celle-ci et en regroupant les textes par type de relations.
Peux-tu regrouper ces textes par famille de relation
en indiquant la nature de la relation pour chaque groupe ?

Merci de donner des exemples uniquement issus du fichier en txt.
```

A.5 [English version] Prompt used for explore relations into administrative documents with LLMs

```
You're an AI assistant tasked with extracting relationships from batches
of texts written by French public administrations.

You receive as input several texts that have been grouped together in the same category,
as well as the named entities present in these texts between brackets,
and you need to determine whether these texts are linked by similar relationships
by indicating the relationship and grouping the texts by relationship type.
Can you group these texts by relationship family,
indicating the nature of the relationship for each group?

Please give examples from the txt file only.
```

A.6 [French version] Prompt used for explore relations into BizRel with GPT-4o

```
Tu es un assistant IA qui est chargé de faire de l'extraction de relations
dans des batch de textes issus du web français.
Tu reçois en entré plusieurs textes ayant été regroupé dans la même catégorie
ainsi que les entités nommées présentent dans ces textes comprise entre
des balises [E11][E12] et [E21][E22], tu dois déterminer si ces textes
sont reliés par des relations similaire en indiquant celle-ci et en
regroupant les textes par type de relations.

Peux-tu regrouper ces textes par famille de relation en indiquant
la nature de la relation pour chaque groupe ?

Merci de donner des exemples uniquement issus du fichier en txt.
```

A.7 Relations generated from our administrative corpus

Labels	Details	Entity pairs
Land or infrastructure acquisition	A public administration acquire land or real estate from another organization or person.	ORG - ORG ORG - PER
Subsidies and financing	Concern the grant for a financing asked or proposed by the administration to another organization.	ORG - ORG
Relations between public institutions	Participation in an event, membership, or the desire to create a project with one or more other administrations. Inter-city sharing of administrative staff.	ORG - ORG
Delegation of public services	Public service delegation to an exterior organization.	ORG - ORG
Management of environmental labels	False positive, example didn't match the category.	None
Located in	Mention of a relation between two geographical entities	LOC - LOC
Loan guarantee	A public authority guarantees a loan for an organization to assist with its economic development.	ORG - ORG
Public procurement and contracts	Awarding a public contract or signing a contract with an organisation.	ORG - ORG
Management of municipal and administrative services	Administrative organization, new public services available to the community.	ORG - ORG
Part of an organisation	An elected official could be part of an organization implying one or multiple public structures.	PER - ORG
Public works and infrastructure	Refurbishment and renovation of public buildings. Work on roads and public spaces.	ORG - LOC
Administrative litigation	False positive relative to reminder of the law concerning administrative appeal procedures. Or legal framework for amending agreements and endorsements to municipal contracts.	None
Vote or political position on the community board	Participation in community board votes: Taking a position for or against it, or taking no position.	PER- PER
Position or mandate held within an organization	Elected officials functions or assignment of temporary responsibilities.	ORG - PER
Collaboration, Discussion, or Mention in an Administrative Context	Collaboration, discussion, or mention of individuals in the context of a meeting, debate, or administrative action.	PER - PER
Administrative Decision or Action	An elected official is mention for give a decision about an organization	PER - ORG

Table 6: Labels and details on the relations generated from administrative data, using sentence clustering by entity pair.