Can LLMs Translate Italy's Language Varieties?

Edoardo Signoroni
NLP Centre, Faculty of Informatics
Masaryk University
Botanická 68a, 602 00 Brno, Czechia

Pavel Rychlý
NLP Centre, Faculty of Informatics
Masaryk University
Botanická 68a, 602 00 Brno, Czechia

Abstract

We evaluate the capabilities of several small large language models (LLMs) to translate between Italian and six low-resource language varieties from Italy (Friulan, Ligurian, Lombard, Sicilian, Sardinian, and Venetian). Using recent benchmark datasets, such as FLORES+ and OLDI-Seed, we compare prompting and fine-tuning approaches for downstream translation, evaluated with CHRF scores. Our findings confirm that these LLMs struggle to translate into and from these low-resource language varieties. Pretraining and fine-tuning a small LLM did not yield improvements over a zero-shot baseline. These results underscore the need for further NLP research on Italy's low-resource language varieties. As the digital divide continues to threaten the conservation of this diverse linguistic landscape, greater engagement with speaker communities to create better and more representative datasets is essential to boost the translation performance of current LLMs.

1 Introduction

Recent advances in Large Language Models (LLMs) have significantly improved multilingual natural language processing, enabling high-quality machine translation and other downstream tasks for many major world languages. However, the benefits of these models are unevenly distributed, with low-resource languages remaining underrepresented in both training data and model capabilities.

Italy is home to a uniquely diverse linguistic landscape, featuring regional and minority language varieties alongside Standard Italian. While Standard Italian achieved official status with the country's unification, it became widespread only after the birth of mass media, with most of the population speaking a local language variety in everyday life. Over time, most local varieties declined, having faced not only marginalization, but also legal ambiguity and social stigma. Even if some scattered interest is re-emerging, even in NLP, more

than 30 language varieties are endangered, as recognized by UNESCO. Without systematic evaluation and targeted development, speakers of Italy's local varieties are excluded from the advantages of modern language technology, contributing to cultural loss.

This work aims to address these challenges by assessing the performance of several LLMs in translating between Italian and six low-resource varieties: Friulan, Ligurian, Lombard, Sicilian, Sardinian, and Venetian. We evaluate both zero-shot, few-shot, continued pretraining, and fine-tuning approaches using recent machine translation benchmark datasets.

Our objective is to measure the translation quality of small sized LLMs for Italy's language varieties, in order to confirm current limitations of LLMs and highlight avenues for future research and resource development.

The remainder of the paper is organized as follows. Section 2 provides an overview of Italy's language varieties and their sociolinguistic context. Section 3 reviews related work in low-resource machine translation, and for the language varieties of Italy. Section 4 details our experimental setup, datasets, and models. Section 5 presents and discusses the results, followed by conclusions and future directions in Section 6.

2 Language Varieties of Italy

Italy presents one of the most diverse language landscapes in Europe (Maiden and Parry, 1997). Standard Italian was adopted as the national language only relatively recently, after the unification in 1861, at a time when it was spoken only by less than 10% of the population, mostly in Rome and Tuscany. In fact, Italian itself was a development of a literary language based on Vulgar Latin and from the Tuscan variety spoken by the Florentine upper class. In the first phases of the nation's his-

tory it was important for the state to standardize the local varieties to ensure the smooth operations of officials and teachers. With illiteracy remaining widespread for most of the 1800s and 1900s, the rise of education and mass media cemented the widespread use of Standard Italian, with detriment to the use of local languages.

During the Fascist era, local and minority languages were suppressed, under the pretext of them undermining central authority. A severe linguistic policy of italianization was enforced throughout the nation and its occupied territories (van der Jeught, 2016; Ramponi, 2024). Local language varieties were banned even in everyday life, teaching was allowed only in Italian, toponyms and even surnames were changed to Italian-sounding forms. For this reason, local language varieties are stigmatized as a sign of ignorance and lack of integration, and denoted with the negative-charged and linguistically improper connotation of the term *dialetti* (dialects), implying a derivative status as "dialects of Standard Italian". As Ramponi (2024) points out, the term *language varieties* is a more politically neutral denomination, preventing judgment on the prestige and status of each language.

The republican constitution of 1946 provides for the protection of linguistic minorities, but the specific implementation, through the Law 482/1999, is demanded to the local level and covers only a handful of language varieties (Albanian, Catalan, Germanic, Greek, Slovenian, Croatian, French, Francoprovençal, Friulian, Ladin, Occitan, and Sardinian). Others are protected by regional laws (Arbëreshë Albanian in Apulia and Calabria; Algherese Catalan, Gallurese, Sardinian, and Sassarese in Sardinia; German in the Walser-speaking Valle del Lys (Aosta Valley); Cimbrian, Ladin, and Mòcheno in Trentino; Calabrian Greek and Occitan (i.e., Vivaro-Alpine Gardiol) in Calabria; Francoprovençal (i.e., Faetar) and Griko in Apulia) or only recognized (Lombard, Piedmontese, and Sicilian in Lombardy, Piedmont, and Sicily, respectively; Promoted: Friulian and Slovenian in Friuli-Venezia Giulia, and Francoprovençal, French, Occitan, and Walser in Piedmont; Both: Venetian in Veneto and Ligurian Tabarchino in Sardinia) (Ramponi, 2024). This also reflects in the current status of the varieties: some are used only by very small communities of elders, while others are co-official with Italian and used in schools and education in their specific region, e.g. German and French. There exist cultural institutes that promote initiatives on language and cul-

ture for small language groups, but most often the non-officially recognized varieties are supported by politically-motivated and polarized groups.

In 2017, the ISTAT (*Istituto Nazionale di Statistica*, National Institute of Statistics) reported that 45.9% of the population mainly speak Italian at home, 32.3% use Italian and a local language, and 14.1% mostly speak a local language (ISTAT, 2017). Most local language varieties live in *dialalìa*[1] (Ramponi, 2024) with Italian, which is used in all formal and official settings, whereas the local variety are more and more confined to informal situations, overlapping with Italian even in these domains. For this reason, even if some historical literary traditions exist, most of the language varieties are primarily used in spoken and informal settings, and lack a codified written form, with the speakers improvising writing "as the words sound". When the speakers write their variety, it is often in code-switching.

Of the many language varieties present in Italy, more than 30 are endangered according to UNESCO (Moseley, 2010; Ramponi, 2024). According to Joshi et al. (2020), 10 endangered varieties are in the second-to-last position of the scale, while the rest are last (Ramponi, 2024). For consistency and ease of access, we limit our exploratory evaluation to the language varieties that appear in the OLDI-Seed (Costa-jussà et al., 2024) dataset, summarized in Table 1.

3 Related Work

In this Section, we first report on related work regarding prompting LLMs for low-resource machine translation (3.1) and then focus on specific machine translation resources for the languages we experiment with (3.2).

3.1 LLMs for Low-Resource Machine Translation

Some recent work explores and leverages LLMs prompting and adaptation for low-resource machine translation. Robinson et al. (2023) evaluates ChatGPT translation on the FLORES-200, the previous version of FLORES+. Thy find that for high-resource languages, it performs on par or ex-

---

[1] This is a particular form of *diglossia*, where there is not a clear functional separation between "high" and "low" varieties. The high variety, e.g. Italian, is used in formal contexts, but also encroaches in everyday informal communication, which in *diglossia* would be reserved only for the "low" variety, e.g. Lombard.

ID	Name	Branch	Conservation	Status	Standard	Speakers
fur	Friulan	Romance (Rhaeto-Romance)	DE	P, p	Y	0.6M
lij	Ligurian	Romance (Gallo-Italic)	DE	p*	Y	0.5M
lmo	Lombard	Romance (Gallo-Italic)	DE	r	N	3.5M
scn	Sicilian	Romance (Italo-Romance)	V	r	N	4.7M
srd	Sardinian	Romance (Sardinian)	DE	P, p**	Y	1.0M
vec	Venetian	Romance (^{o})	V	p	Y	3.9M

Table 1: Table of the language varieties in our experiments. For each one, the table reports its ISO-639-3 code, its name, its genealogy, the conservation status according to Moseley (2010) (Definitely Endangered [DE], Vulnerable [V], its status (Protected [P] varieties are those so by national law, whereas Recognition [r] and Promotion [p] are at regional level. For these varieties, Promotion implies recognition), if an official orthography exists, and the number of speakers (Ramponi, 2024). *Only *Tabarchino* in Sardinia **In its Gallurese and Sassarese varieties. ^{o}Disputed, either Gallo-Italic, or Italo-Dalmatian

ceeds traditional MT models, but struggles for low-resource languages, confirming that a language's resource level is the most important feature in determining ChatGPT's relative ability to translate it.

Shu et al. (2024) employ a keyword translation and retrieval method to augment translation for low-resource languages. They test their approach with GPT-4o and LLaMA 3.1 405B, which struggle in a zero-shot low-resource scenario, and outperform the baselines.

Merx et al. (2024) test LlaMA 2 70b, Mixtral 8x7B, and GPT-4 on English-Mambai translation. They explore few-shot prompting with a novel corpus prepared by a language manual and supplemented with sentences from a native speaker. They include dictionary entries, sentences, and semantic embeddings in the prompts significantly improves the translation. However, their experiments also show wide fluctuations in BLEU score across different domains.

Guo et al. (2024) devise a way to teach LLMs to translate low-resource languages by guiding them with a textbook-like approach. They assess this method on FLORES+ with ChatGPT and BLOOMZ and achieve better performance than zero-shot baselines by enhancing the models' knowledge to generate accurate and fluent sentences.

Tanzer et al. (2024) introduce a method for learning English-Kalamang translation using several pages of field linguistics reference materials, thus enabling a model to learn from a human-readable grammar book. While the experiments are promising, they still fall short of human performance. Hus and Anastasopoulos (2024) reuse the same method, leveraging the long context of GPT-4 and

improve the performance of machine translation for 16 low-resource languages. Aycock et al. (2025) however, deconstruct these experiments on Nepali and Guarani, showing that the biggest contribution is from task relevant data in the book, that is parallel examples for translation, and grammatical information for linguistic tasks.

3.2 Machine Translation for the Varieties of Italy

Some work was done for the language varieties of Italy. Here we focus on datasets and tools related to machine translation for the languages we cover in this paper. For a wider survey of NLP for the language varieties of Italy, we direct the reader to Ramponi (2024).

Delmonte et al. (2009) investigate Venetian-English machine translation, between other NLP tools needed to train their statistical model, such as a PoS tagger and a rule-based system to translate Italian text into Venetian.

Wdowiak (2022) develop a Sicilian-Italian and Italian-Sicilian neural machine translation system using small subword vocabularies to train Transformers models with a high dropout. Using further techniques, such as backtranslation and multilingual translation, and the incorporation of theoretical knowledge, they manage to reach BLEU scores in the 30s.

Tyers et al. (2017) create the first machine translation system from Italian to Sardinian, using a rule-based approach. Their system achieves the stated objective of generating Sardinian text ready for post-editing. Fronteddu et al. (2017) presents a Catalan-Sardinian translator.

Signoroni (2022) describes a human-evaluated, revised, and corrected Lombard-Italian parallel cor-

pus destined to train machine translation systems. With the help of bilingual annotators, they audit an automatically aligned Wikipedia corpus from OPUS (Tiedemann, 2009).

Haberland et al. (2024) presents a Ligurian-Italian (and partially English) parallel corpus, specifically tailored for the cultural environment of Ligurian speakers. They find that using this corpus improves the performance of translation models compared to NLLB-3.3B.

Multilingual language models such as mBERT (Devlin et al., 2019) includes a tiny percentage of data from Lombard, Piedmontese, and Sicilian. However, these data are often indiscriminately collected from Wikipedia, regardless of linguistic quality or representativeness.

4 Methodology

In the following Section, we first describe the datasets we use (4.1) and then the methodology of our experiments (4.2).

4.1 Data

In our experiments, we use FLORES+ (dev and devtest) and OLDI-Seed (train) (Costa-jussà et al., 2024). Both benchmark datasets comprise professionally translated sentences (1k for FLORES+ and 6k for OLDI-Seed) from Wikipedia. In our experiment on pretraining (see Section 4.2), we use the target portion of the OLDI-Seed as monolingual pretraining data. For multilingual pretraining, we combine the data for all languages.

4.2 Experiments

Our goal is to evaluate the performance of existing small-to-mid sized LLMs when prompted for translation between Italian and several local languages of Italy. We chose to translate to and from Italian because this would be the most probable real-life application of this technology.

Prompting For prompting, we use Ollama.[2] We prompt three instruct models, Mistral NeMo (Mistral-AI, 2024), EuroLLM-9B-Instruct (Martins et al., 2024), and Qwen2.5-14B-Instruct (Qwen et al., 2025); and two reasoning models, Deepseek-R1-14B (DeepSeek-AI et al., 2025) and Phi-4-Reasoning (Abdin et al., 2024). We experiment with 0-shot, 1-shot, and 3-shots prompts both in English and Italian. Examples for the few-shot

prompts are sampled from the *dev* split of FLORES+. We use default setting, but set the temperature at 0.7.

Pretraining and Fine-Tuning To investigate the effects of continued pretraining and fine-tuning fully, we limit our experiments to a small LLM, Qwen2.5-05.B-Instruct (Qwen et al., 2025), due to hardware and compute constraints. We test several pretraining and fine-tuning combinations, summarized in Table 2. We then prompt these models for translation into the low-resource varieties. For both experimental phases, we use the LitGPT library.[3] We use a local server with a combination of NVIDIA A40 and A100. We pretrain the models on a single GPU, for 10 epochs with an learning rate of 1e-4 and a batch size of 2048 tokens. For fine-tuning, we train the models for 3 epochs, with a learning rate of 1e-4, and a batch size of 16.

To evaluate the results we compute BLEU (Papineni et al., 2002), CHRF (Popović, 2015), and Comet (Rei et al., 2020) with HuggingFace `evaluate` library. Due to the low-resource nature of the language in the experiment, we prefer CHRF over the other metrics, since Comet is not reliable for unseen languages.

5 Results and Discussion

Below, we discuss the results for prompting (5.1 and fine-tuning experiments (5.2).

5.1 Prompting

General Observations As expected, translation into Italian is ∼10-18 CHRF better than the other direction. Less expected is the sub-par performance of the *reasoning* models, trailing behind of an average ∼13 CHRF points. For both type of models and directions, *Venetian* was the easiest language for the models.

Instruct Models Among the *instruct* models, `eurollm-9b-instruct` (eurollm) is the best performing for each language and direction. Interestingly, when using English prompts, it loses performance the more shots are given (39.6 > 37.5 CHRF). Conversely, with an Italian prompt, the model retains quality, but does not improve with more shots (40.2-5 CHRF). Using more shots for `mistral-nemo:12b` (mistral) degrades its performance with both prompts, from ∼39 to ∼36 CHRF.

Pretrain	Pretrain Data	Fine-Tune	Fine-Tune Data	Name
N	-	N	-	no-pt_no-ft
Y	monolingual	N	-	mono-pretrain
Y	multilingual	N	-	multi-pretrain
N	-	full	monodirectional	full-mono
Y	multilingual	full	monodirectional	multi-pretrain_full-mono
Y	monolingual	full	monodirectional	mono-pretrain_full-mono
Y	multilingual	full	multidirectional	multi-pretrain_full-multi
N	-	qlora	monodirectional	qlora-mono

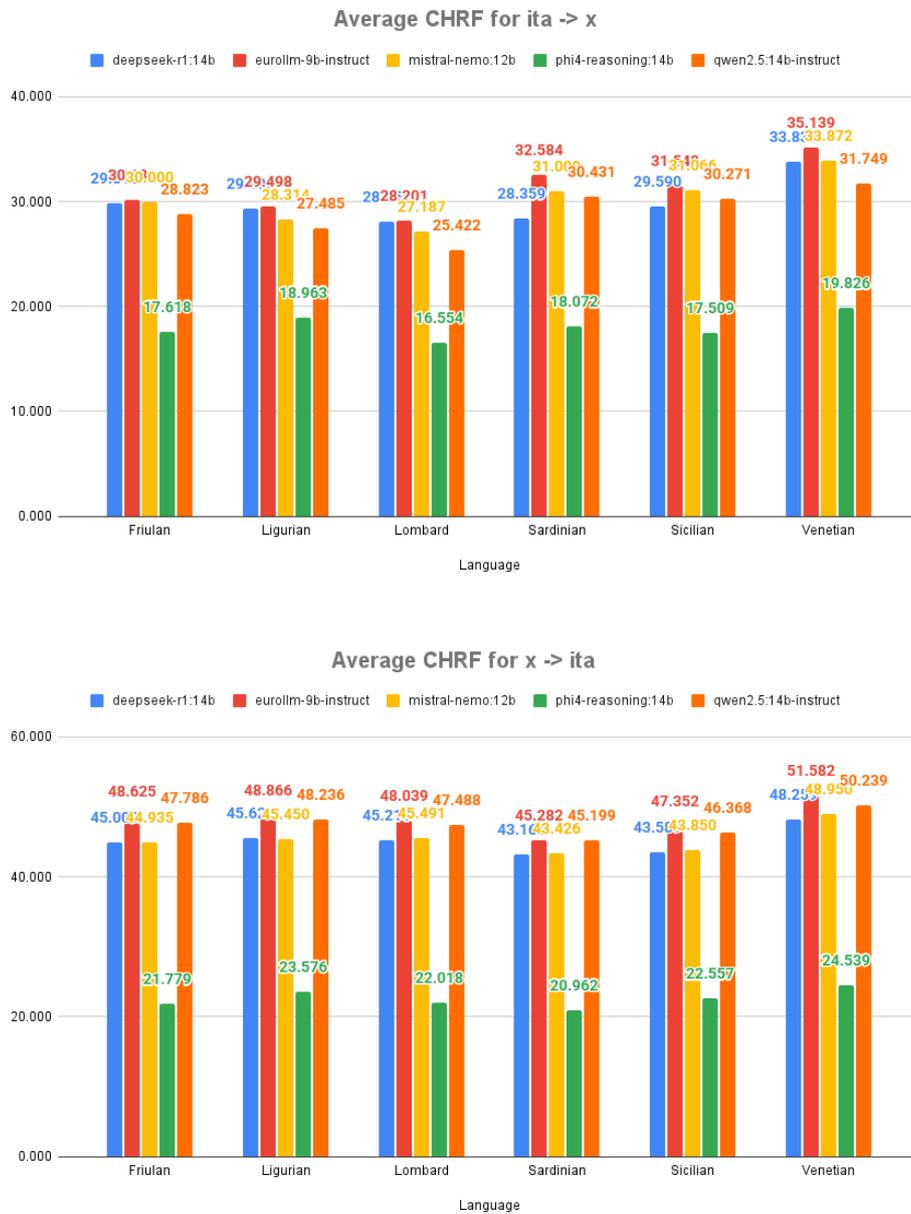Table 2: Combinations of pretraining and fine-tuning for `qwen2.5:14b-instruct`.

Figure 1: Average CHRF score for each model and language, when translating from and into Italian.

Conversely, `qwen2.5:14b-instruct` (qwen) benefits from few-shot prompts, from ∼37 to ∼39 CHRF. For the *instruct* models, Lombard is the hardest language to translate into. When translating into Italian, Sardinian is the harder to process.

Reasoning Models *Reasoning* models did not perform better than *instruct* ones. `deepseek-r1:14b` (deepseek) performs comparably to `mistral` and `phi4-reasoning:14b` (phi) is the worst performing model overall, with roughly half of the CHRF score of the other models (∼20 CHRF on average). deepseek's performance stays stable with few-shot and Italian prompts, while phi get worse the more examples it is given. Using Italian prompts does not change its behavior.

5.2 Pretraining and Fine-Tuning

Among all the tested models, zero-shotting `qwen2.5:0.5b-instruct` gives the best results. Despite the huge difference in size, qwen performs almost as well as the other models for most of the languages, and even reaching `mistral` and deepseek on Lombard and Venetian. Just behind the baselines, we find *mono-pretrain*, that is continued pretraining with monolingual target data. Future work is warranted to see if more and more diverse data could push the model above the baseline for all languages, not only for Friulan. Full finetuning with monodirectional data is effective only for Lombard, and achieves worse performance than pretraining with other directions. When paired with pretraining, full finetuning is detrimental, as is multilingual and multidirectional data. QLoRA performs the worst overall, with `mono-pretrain_full-mono` under-performing it only for Ligurian and Venetian.

6 Conclusions

Italy has a rich and diverse linguistic landscape. However, most of its language varieties are endangered and lack modern NLP resources and tools. LLMs are being leveraged to fill the gap between high-resource and low-resource languages also for machine translation. We tested six such models for six language varieties of Italy. Our results show that LLMs alone are not enough to close the digital divide between Italian and local low-resource language varieties. These unseen languages, unlike high-resourced Italian, pose a challenge for both instruct and newer reasoning small-to-mid-sized models, leading to poor translation performance.

Comparison with a smaller 0.5B model shows that bigger is not always better. Regardless of size, LLMs cannot translate what they do not see, and more high-quality monolingual data could ameliorate the issue.

Limitations

Our work is intended as a starting point for work on Italy's underresourced language variants. We do not engage in extensive parameter tuning, nor in data collection or augmentation to improve the quality of the translation. These tasks are often language-specific and thus are left to future work.

Another limitation has to do with the data. Due to the complex landscape of Italy's language varieties, the data for at least some of the languages involved, as acknowledged also by Costa-jussà et al. (2024), may not be representative of all speakers of the state variety or language, such as the case of Lombard, for example, which does not have a standardized orthography. Moreover, the data we used covered only a small fraction of the language varieties present in Italy, and only in some specific form that may or may not be completely mirrored in everyday usage.

For the same reasons, a deep qualitative evaluation was not possible due to widespread differences in orthography and lexicon.

Ethics Statement

Prompting and fine-tuning LLMs is an energy- and data-hungry endeavor. We estimate the carbon footprint of our experiments at 98.82 kgCO$_2$eq (Lacoste et al., 2019).

Although we do not intend to deploy the systems we trained for actual, real-word usage, we note that in such cases they may produce unreliable, biased, incorrect, or not representative output, also for the complex language issues mentioned in Sections 2 and 6.
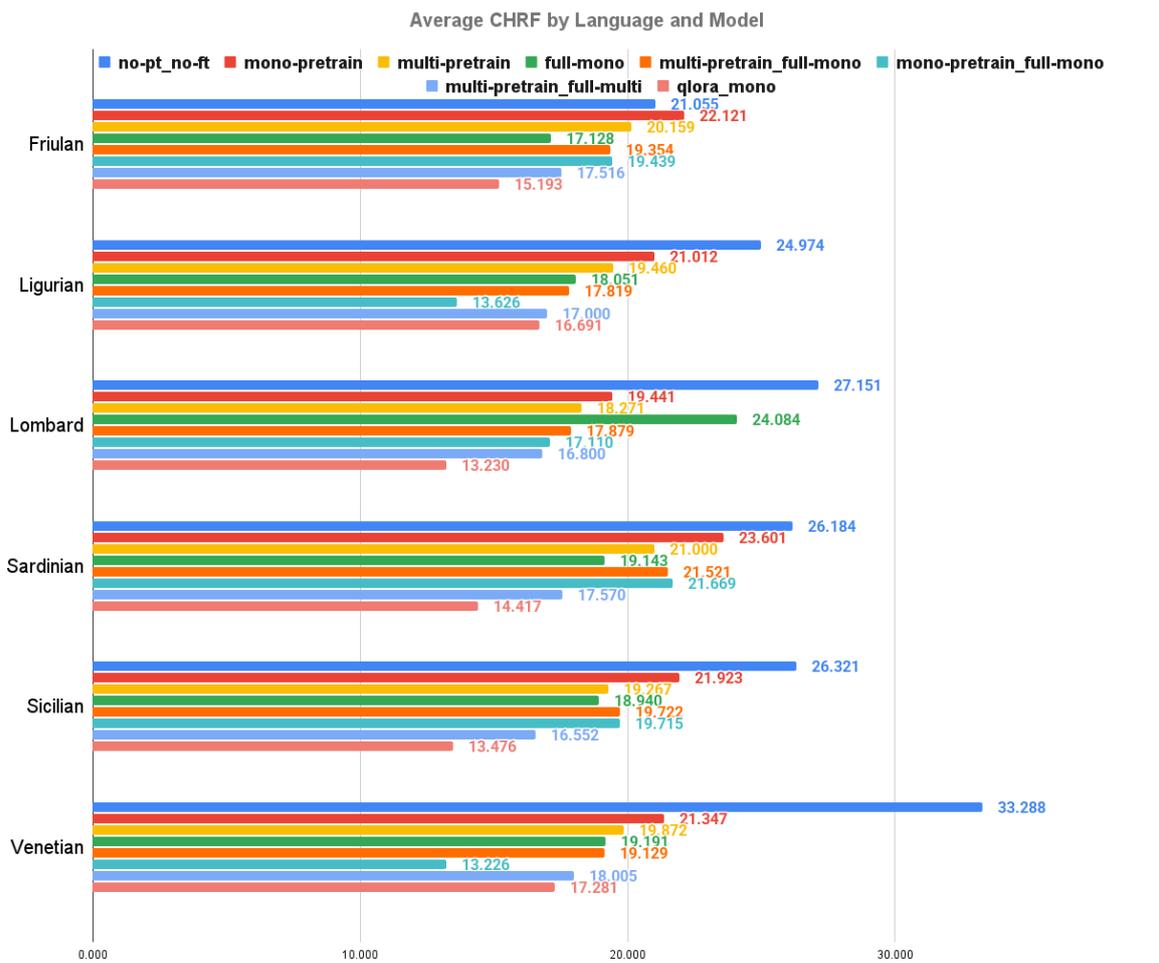
Figure 2: Average CHRF for pretrained and fine-tuned qwen2.5:0.5b-instruct models. *no-pt_no-ft* is the baseline model. All the models were prompted without additional examples.

References

Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J. Hewett, Mojan Javaheripi, Piero Kauffmann, James R. Lee, Yin Tat Lee, Yuanzhi Li, Weishung Liu, Caio C. T. Mendes, Anh Nguyen, Eric Price, Gustavo de Rosa, Olli Saarikivi, and 8 others. 2024. Phi-4 technical report. *Preprint*, arXiv:2412.08905.

Seth Aycock, David Stap, Di Wu, Christof Monz, and Khalil Sima'an. 2025. Can LLMs really learn to translate a low-resource language from one grammar book? In *The Thirteenth International Conference on Learning Representations*.

Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, and 20 others. 2024. Scaling neural machine translation to 200 languages. *Nature*, 630(8018):841–846.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, and 181 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

Rodolfo Delmonte, Antonella Bristot, Sara Tonelli, and Emanuele Pianta. 2009. English/veneto resource poor machine translation with stilven. In *Proceedings of ISMTCL*, pages 82–89, Besançon, France. Presses Universitaires de Franche-Comté.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Gianfranco Fronteddu, Hèctor Alòs i Font, and Francis M. Tyers. 2017. Una eina per a una llengua en procés d'estandardització: El traductor automàtic català-sard. *Linguamática*, 9(2):3–20.

Ping Guo, Yubing Ren, Yue Hu, Yunpeng Li, Jiarui Zhang, Xingsheng Zhang, and Heyan Huang. 2024. Teaching large language models to translate on low-resource languages with textbook prompting. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15685–15697, Torino, Italia. ELRA and ICCL.

Christopher R. Haberland, Jean Maillard, and Stefano Lusito. 2024. Italian-Ligurian machine translation in its cultural context. In *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024*, pages 168–176, Torino, Italia. ELRA and ICCL.

Jonathan Hus and Antonios Anastasopoulos. 2024. Back to school: Translation using grammar books. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20207–20219, Miami, Florida, USA. Association for Computational Linguistics.

ISTAT. 2017. L'uso della lingua italiana, dei dialetti e di altre lingue in italia. https://www.istat.it/it/archivio/207961. Accessed: 2025-08-07.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. 2019. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*.

Martin Maiden and Mair Parry. 1997. *The Dialects of Italy*. Routledge.

Pedro Henrique Martins, Patrick Fernandes, João Alves, Nuno M. Guerreiro, Ricardo Rei, Duarte M. Alves, José Pombal, Amin Farajian, Manuel Faysse, Mateusz Klimaszewski, Pierre Colombo, Barry Haddow, José G. C. de Souza, Alexandra Birch, and André F. T. Martins. 2024. Eurollm: Multilingual language models for europe. *Preprint*, arXiv:2409.16235.

Raphaël Merx, Aso Mahmudi, Katrina Langford, Leo Alberto de Araujo, and Ekaterina Vylomova. 2024. Low-resource machine translation through retrieval-augmented LLM prompting: A study on the Mambai language. In *Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI) @ LREC-COLING 2024*, pages 1–11, Torino, Italia. ELRA and ICCL.

Mistral-AI. 2024. Mistral nemo. https://mistral.ai/news/mistral-nemo. Accessed: 2025-08-07.

Christopher Moseley. 2010. *Atlas of the World's Languages in Danger*. UNESCO Publishing.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, and 25 others. 2025. Qwen2.5 technical report. *Preprint*, arXiv:2412.15115.

Alan Ramponi. 2024. Language varieties of Italy: Technology challenges and opportunities. *Transactions of the Association for Computational Linguistics*, 12:19–38.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.

Nathaniel Robinson, Perez Ogayo, David R. Mortensen, and Graham Neubig. 2023. ChatGPT MT: Competitive for high- (but not low-) resource languages. In *Proceedings of the Eighth Conference on Machine Translation*, pages 392–418, Singapore. Association for Computational Linguistics.

Peng Shu, Junhao Chen, Zhengliang Liu, Hui Wang, Zihao Wu, Tianyang Zhong, Yiwei Li, Huaqin Zhao, Hanqi Jiang, Yi Pan, Yifan Zhou, Constance Owl, Xiaoming Zhai, Ninghao Liu, Claudio Saunt, and Tianming Liu. 2024. Transcending language boundaries: Harnessing llms for low-resource language translation. *Preprint*, arXiv:2411.11295.

Edoardo Signoroni. 2022. Piötòst ché nìent, mèi piötòst-a manually revised lombard-italian parallel corpus. *RASLAN 2022 Recent Advances in Slavonic Natural Language Processing*, page 105.

Garrett Tanzer, Mirac Suzgun, Eline Visser, Dan Jurafsky, and Luke Melas-Kyriazi. 2024. A benchmark for learning to translate a new language from one grammar book. *Preprint*, arXiv:2309.16575.

Jörg Tiedemann. 2009. *News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces*, volume V, pages 237–248.

Francis M. Tyers, Hèctor Alòs i Font, Gianfranco Fronteddu, and Adrià Martín-Mor. 2017. Rule-based machine translation for the italian–sardinian language pair. *The Prague Bulletin of Mathematical Linguistics*, 108:221–232.

Stefaan van der Jeught. 2016. The protection of linguistic minorities in italy: A clean break with the past. *Journal on Ethnopolitics and Minority Issues in Europe*.

Eryk Wdowiak. 2022. A recipe for low-resource nmt. In *Intelligent Computing (SAI 2022)*, pages 739–746, Cham, Switzerland. Springer International Publishing.