The *Mediomatix* Corpus: Parallel Data for Romansh Language Varieties via Comparable Schoolbooks

Zachary Hopton[1] **Jannis Vamvas**[1] **Andrin Büchler**[2]
Anna Rutkiewicz[1] **Rico Cathomas**[2] **Rico Sennrich**[1]
[1]University of Zurich [2]University of Teacher Education of the Grisons
zacharywilliam.hopton@uzh.ch, vamvas@cl.uzh.ch

Abstract

The five idioms (i.e., varieties) of the Romansh language are largely standardized and are taught in the schools of the respective communities in Switzerland. In this paper, we present the first parallel corpus of Romansh idioms. The corpus is based on 291 schoolbook volumes, which are comparable in content for the five idioms. We use automatic alignment methods to extract 207k multi-parallel segments from the books, with more than 2M tokens in total. A small-scale human evaluation confirms that the segments are highly parallel, making the dataset suitable for NLP applications such as machine translation between Romansh idioms. We release the parallel and unaligned versions of the dataset under a CC-BY-NC-SA license[1] and demonstrate its utility for machine translation by training and evaluating an LLM and a supervised multilingual MT model on the dataset.

1 Introduction

Romansh, a Romance language spoken in southeast Switzerland, has five main regional varieties, which are commonly called *idioms*. Even though the variability of Romansh idioms takes a central role in the Romansh language community, including in school education, there has been no parallel corpus to date, limiting the prospect of machine translation (MT) between Romansh idioms.

In this paper, we present the first (multi-)parallel corpus for the five Romansh idioms, which we extracted from 291 comparable schoolbooks of the *Mediomatix* series for teaching Romansh as a first language. Figure 1 presents an example of a short parallel segment across the five idioms, illustrating the variance between the idioms in terms of lexicon, syntax, and orthography. In total, our multi-parallel corpus contains 207k aligned segments of varying lengths, spanning more than 2 million tokens. We release it to the research community with permission from the schoolbook editors.

Sursilvan:	*Speronza ei quei ver.*
Sutsilvan:	*Sprànza e quegl ver.*
Surmiran:	*Speranza è chegl veir!*
Puter:	*Spraunza es que vaira!*
Vallader:	*Spranza es quai vaira!*
[English]	*Hopefully that's true!*

Figure 1: Example of a parallel segment in the five Romansh idioms.

We automatically align the segments by using VecAlign (Thompson and Koehn, 2019), a standard approach for embedding-based parallel sentence alignment. To enable a high-precision multi-parallel alignment across the five idioms, we use an approach that we call *pivot consensus alignment*. We validate the alignment (1) by reporting accuracy on a validation set, and (2) by performing a small-scale evaluation through a native Romansh speaker, which showed that 471 out of 472 evaluated segments were aligned correctly.

We further demonstrate that our corpus can be successfully used for fine-tuning NLP systems on the task of MT from one idiom into another, which, to our knowledge, is the first such attempt for Romansh idioms. Code to reproduce our experiments is available.[2]

2 Background

2.1 Language Situation

Alongside German, French, and Italian, Romansh is one of Switzerland's four national languages (ISO 639-1: `rm`; ISO 639-2/3: `roh`). It has an estimated number of 60,000 speakers (Müller and

[1] Parallel Corpus: `https://huggingface.co/datasets/ZurichNLP/mediomatix`; Unaligned Corpus: `https://huggingface.co/datasets/ZurichNLP/mediomatix-raw`

[2] `https://github.com/ZurichNLP/mediomatix-code`

Findings of the Association for Computational Linguistics: EACL 2026, pages 290–306
March 24-29, 2026 ©2026 Association for Computational Linguistics

Idiom	**Book volumes**	**Segments**		**Tokens**	
	(workbook + commentary)	Overall	Aligned	Overall	Aligned
Sursilvan	67	104,481	49,872	955,606	513,036
Sutsilvan	67	104,825	49,137	989,282	513,850
Surmiran	27	51,723	14,301	463,694	145,982
Puter	65	106,348	47,185	993,705	494,850
Vallader	65	113,754	47,397	1,058,343	495,404
Total	**291**	**481,131**	**207,892**	**4,460,630**	**2,163,122**

Table 1: Dataset statistics for the *Mediomatix* corpus.

Roth, 2019) and enjoys legally protected minority status (Etter, 2018). Romansh's present-day situation is strongly influenced by its unique sociolinguistic context (Grünert, 2024). The "traditional" speaking area in the canton of Grisons comprises five core territories which are divided geographically but also linguistically.

Romansh Idioms Each of these territories has its "own" Romansh idiom (i.e., regional standard variety). These idioms are **Sursilvan** (estimated 55% of Romansh speakers) in western Grisons, **Sutsilvan** (estimated 3% of Romansh speakers) and **Surmiran** (estimated 10% of Romansh speakers) in central Grisons, **Puter** (estimated 12% of Romansh speakers) in the upper Engadine valley, and **Vallader** (estimated 20% of speakers) in the lower Engadine valley (Furer, 2005). Each of the Romansh idioms are standardized, written forms of the language that have emerged over the last 400 years (Caviezel, 1993; Liver, 2010) and which have their own codices (e.g., Spescha, 1989) and literary traditions. There are therefore substantial differences not only in orthography but also other linguistic levels such as vocabulary (e.g., *glimaglia* in Sursilvan vs. *lindorna* in Vallader for 'snail') or morphosyntax (e.g., analytic future tense using the auxiliary verb *to come* in Sursilvan and Sutsilvan vs. synthetic future tense in the other idioms). As a result, mutual intelligibility is sometimes challenging for speakers of differing idioms.

Contemporary Role of Romansh Idioms Evolution of five differing regional standard varieties was fostered by factors such as processes of demarcation to avoid intermixing of idioms (Arquint, 1982) or strong regional attachment and closeness to the vernacular (Diekman, 1991). In the 1980s, linguists and policy makers created and implemented a supraregional standard variety (Cathomas, 2024; Coray, 2010; Schmid, 1989). This variety, known as *Rumantsch Grischun*, entails elements from all idioms (Gross, 1999). Nowadays, Rumantsch Grischun is used partly in media, in higher education, and in federal and cantonal administration (Grünert, 2024) but only in a very limited number of schools as language of instruction. Consequently, the broader population is not literate in Rumantsch Grischun but in one of the five idioms.

2.2 NLP for Romansh

While Romansh idioms have been studied in the context of commercial speech technology[3], previous work on the processing of Romansh text has focused on Rumantsch Grischun, the supraregional standard form of the language (e.g., Müller et al., 2020; Dolev, 2023; Vamvas et al., 2023). With the *Mediomatix* corpus, we hope to provide a basis for multilingual NLP research on written text in the five regional idioms of Romansh.

2.3 Teaching Materials as a Source of Parallel Text for Romansh Idioms

The five Romansh idioms are used as school languages in their respective areas, alongside German as an equal or dominant language of instruction. Additionally, Romansh is explicitly taught as a subject (Cathomas, 2005; Gross, 2017). For the latter case, a series of teaching materials is being developed for grades 2 to 9 (i.e., age 8 to 16) at the University of Teacher Education of the Grisons, called *Mediomatix*.[4]

Once complete, *Mediomatix* will comprise 325 components for teaching and learning: 160 workbooks (32 per idiom), 160 commentaries (32 per idiom) and five grammar books. In total, this sums to 16,000 pages containing around 80,000 structural

[3] https://recapp.ch/

[4] https://mediomatix.ch/

elements (texts, tasks, exercises, images, instructions, charts, links, etc.). This large compilation of textual data in different idioms is representative of Romansh school or academic language (i.e., more formal than everyday speech). Because several language experts are involved in writing and proofreading, the material is highly compliant with the language norms found in the idioms' codices.

The completion of *Mediomatix* is scheduled for 2029, but more than 150 volumes are already available (Appendix J), which we believe to be a suitable basis for compiling a well-controlled, near-parallel corpus of Romansh idioms.

3 Dataset Creation

We create the corpus with standard NLP tools, dividing the process into two steps: (1) text extraction and segmentation; (2) embedding-based alignment of the segments with Vecalign (Thompson and Koehn, 2019), using a *pivot consensus alignment* strategy to create a consistent multi-parallel alignment across all five idioms.

3.1 Extraction of Text Segments

We extract the text from the content management system used for editing the schoolbooks, so OCR is not necessary. Among the extracted content, we retain the original content with all HTML markup, as well as the plain text extracted from the HTML. For splitting the text into segments we follow the HTML markup, treating every paragraph, list item, etc. as a separate segment, and do not perform further sentence splitting.

3.2 Multi-parallel Alignment

Our goal is to create a multi-parallel corpus, where each segment in one idiom is aligned with its corresponding segment in the other four idioms, such as in Figure 1. To reduce complexity, we break down multi-parallel alignment into a series of pairwise alignments, which we then aggregate into a single multi-parallel alignment.

Bilingual Alignments We manually align chapters within each schoolbook volume based on their titles, and then perform automatic bilingual segment alignment within each of the aligned chapters. We do so using the Vecalign algorithm, as previous work shows its effectiveness in low-resource settings that are near-monotonic (Signoroni and Rychlý, 2023). Using the "maximum alignment size" hyperparameter of Vecalign, we limit the algorithm to outputting just 1–1 alignments and deletions (i.e., 1–0 and 0–1) to increase the precision of the alignment.

Choice of Embedding Model We manually construct a multi-parallel validation set of 150 rows to help us choose a suitable strategy for creating segment embeddings. The manual alignment is done by an author using the PDF schoolbooks and a multilingual, multi-idiom Romansh dictionary[5] as references. This allowed us to use formatting cues and English translations to cleanly align the validation set. Using this validation set, we experiment with several embedding models, both open-source and commercial (see Appendix A for experiment details and Appendix B for details about the validation set). Based on the validation results, we decide to use Cohere's `embed-v4.0`[6] to align the full corpus.

Pivot Consensus Alignment To combine the bilingual alignments into a single, multi-parallel alignment, we use a *pivot* idiom (Figure 2). Let i and j be two idioms, and p a pivot idiom. The *pivot alignment* is the set of segments in i and j that are aligned via p:
$A_{ij}^{(p)} = \{(s_i, s_j) | (s_i, s_p) \in A_{ip} \wedge (s_p, s_j) \in A_{pj}\}$. We include pivot-side deletions by adding unmatched segments from A_{ip} and A_{pj} to $A_{ij}^{(p)}$ with null counterparts. The multi-parallel alignment is then the union over all i, j pairs: $\mathcal{A}^{(p)} = \bigcup_{i,j \in \text{idioms}} A_{ij}^{(p)}$. See Appendix C for an example visualization of multi-parallel alignment via a pivot idiom.

Experiments on the validation set (Table 2) indicate that pivot alignments have high recall, e.g., 99.2 when using Sursilvan as the pivot language. However, we want to give a higher weight to precision than to recall in order to minimize the number of misaligned segments in the final corpus. We find that precision can be increased by aggregating the five pivot alignments into a single multi-parallel alignment, using a consensus-based approach.

Specifically, we calculate the pivot consensus alignment as the intersection of the five pivot alignments: $\mathcal{A}^{\text{consensus}} = \bigcap_{p \in \text{idioms}} \mathcal{A}^{p}$.

Length Heuristic Similar to Ng et al. (2019), we filter out segments with mismatching lengths. Segments that are 1.5 times longer or 0.67 times shorter

[5] `https://www.mypledari.ch/`

[6] `https://docs.cohere.com/docs/cohere-embed`

Approach	Prec.	Rec.	F1
Pivoting via Sursilvan only	94.8	99.2	96.9
Consensus across all pivots	97.2	94.1	95.4

Table 2: Taking the consensus across all five possible pivot languages increases precision on the validation set, compared to using a single, arbitrary pivot language.

than the average length in a row are removed from that row.

4 Validation of Alignment Quality

4.1 Accuracy on Validation Set

Table 2 reports precision, recall, and F1-score of the alignment on our validation set. The results show that taking the consensus across all five possible pivot languages yields a higher-precision alignment than using a single pivot language. Since our goal is to maximize precision, we use the consensus alignment for creating the final parallel corpus. Detailed validation results for all pivot languages are provided in Appendix D.

4.2 Human Evaluation of Precision

To evaluate the precision of the final aligned corpus, we randomly select 100 rows from the test set of the corpus. The sampled rows contain 472 segments. We ask a native speaker of Romansh to assess the sample as follows:

- If a segment is notably different from the others in the row (e.g., contains less or more information), but is still generally aligned, it should be marked as noise in parallel data.
- If a segment is misaligned, it should be marked as an alignment error.

Human evaluation shows that 471 out of 472 segments are correctly aligned. Of the correctly aligned segments, 20 are marked as containing noise. The noise occurs within 11 rows, meaning that 89% of the multi-parallel rows are found to be free of noise. Given that manually evaluated translations in several web-crawled parallel corpora contain 50–83% correct translations when accounting for each language pair's data size (Kreutzer et al., 2022), our parallel corpus is relatively high quality. We provide the evaluator instructions in Appendix I, and examples of the evaluator's qualitative feedback in Appendix F.

5 Machine Translation Experiment

We demonstrate that the *Mediomatix* corpus can be used to train and evaluate machine translation between Romansh idioms.

Models We evaluate a version of NLLB-200-1.3B (Costa-jussà et al., 2024) that we fine-tuned on the training split of *Mediomatix*. In addition, we report results for two commercial LLMs, GPT-4o and GPT-4o-mini (OpenAI et al., 2024), in a lower data regime, with either few-shot prompting or fine-tuning on 5000 examples (250 per translation direction).

LLM Prompting For prompting GPT-4o and GPT-4o-mini, we use a similar setup as the WMT24 General Machine Translation Shared Task (Kocmi et al., 2024).[7] We provide the LLMs with 3-shot prompts randomly retrieved from the validation set. The target idiom is specified in natural language (e.g., *"Translate ... into Sursilvan."*); see Appendix G for the full prompt. We generate the LLM translations with greedy decoding.

Fine-tuning The models were trained in a multilingual fashion, i.e., a single model instance was trained jointly on the 20 translation directions.

For fine-tuning NLLB, we use the full training split of *Mediomatix*, as described in Appendix E. We add special tokens for each of the five idioms, which we initialize randomly. Inputs are truncated to 128 tokens. We train with a peak learning rate of 2e-4 and a batch size of 1,500 for 6 epochs, with early stopping based on validation BLEU. For decoding we use beam search with size 4.

For fine-tuning GPT-4o-mini, we use default settings recommended by OpenAI (3 epochs, batch size 10, lr multiplier 1.8).

Evaluation Metric Due to a lack of support of Romansh in trained metrics such as COMET, we report BLEU (Papineni et al., 2002). Future work could collect human judgments of translation quality and adapt trained metrics to Romansh idioms.

Results Table 3 shows SacreBLEU scores (Papineni et al., 2002; Post, 2018)[8] on a subsample of 500 test examples per translation direction[9]. BLEU

[7] `https://github.com/wmt-conference/wmt-collect-translations`

[8] Signature: `#:1|c:mixed|e:no|tok:13a|s:exp|v:2.5`

[9] `https://github.com/ZurichNLP/mediomatix-code/tree/main/mt_experiment/testset_mediomatix`

System	Sursilvan		Sutsilvan		Surmiran		Puter		Vallader		Avg.
	from	into	from	into	from	into	from	into	from	into	
NLLB-200-1.3B (fine-tuned)	50.0	50.8	50.1	49.5	47.6	44.6	53.0	54.3	53.2	54.7	50.8
GPT-4o (few-shot)	31.0	49.3	38.0	21.8	36.8	25.5	38.7	36.9	35.2	46.1	35.9
GPT-4o-mini (few-shot)	27.5	31.8	29.5	26.2	28.7	26.0	33.3	32.9	32.9	34.9	30.4
GPT-4o-mini (fine-tuned)	35.6	39.4	37.5	32.1	34.2	34.0	41.2	41.4	41.0	42.6	37.9

Table 3: Performance of systems translating between Romansh idioms. We report BLEU on a subsample of the *Mediomatix* test split. NLLB-200-1.3B and GPT-4o-mini are fine-tuned with training data from *Mediomatix*. Note that we use different amounts of fine-tuning data: 182,148 sentence pairs for NLLB and 5,000 for GPT-4o-mini.

scores are macro-averaged across all source idioms (for "from") and target idioms (for "into") that are not identical to the given idiom. See Appendix H for BLEU scores on the same subset broken down by source- and target-idiom. Fine-tuning on a subsample of *Mediomatix* improves the ability of GPT-4o-mini to translate between Romansh idioms, with an average improvement of 7.5 BLEU points over the baseline model, and 2 BLEU points over the larger GPT-4o model. The performance of an NLLB model fine-tuned on the full training split of *Mediomatix* is much higher (+12.9 BLEU), demonstrating the benefit of the large *Mediomatix* corpus for enabling MT between Romansh idioms.

6 Conclusion

The *Mediomatix* corpus is an opportunistic, but relatively large, multi-parallel corpus of text in the five Romansh idioms. For the first time, this corpus allows for the training and evaluation of MT systems that translate between the idioms of the Romansh language. Beyond MT systems for end users, this work will allow for new approaches to data augmentation, expanding the availability of other NLP technology in the idioms of Romansh.

Limitations

The corpus creation described in this paper focuses on optimizing precision as opposed to recall, and as a result, a part of the segments in the *Mediomatix* schoolbooks go unused in the aligned corpus (Table 1). Correspondingly, the human evaluation we perform is limited to an evaluation of precision, to make sure that the segments included in the final corpus are indeed aligned correctly.

A second limitation of this resource is that it currently released under a non-commercial license only. We are working with the copyright holders to release the corpus under a more permissive license.

The *Mediomatix* schoolbooks for the different idioms were released at different times. The authors of the books report that content in the later additions to the series (i.e., the Surmiran books) is often translated from the earlier additions, meaning the *Mediomatix* dataset contains so-called "translationese" to some extent. While this may limit the utility of the corpus for MT (Zhang and Toral, 2019), other corpora consisting in translated text like FLORES have found widespread use (Goyal et al., 2022).

Author Contributions

ZH: Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.
JV: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing.
AB: Writing – original draft, Writing – review & editing.
AR: Data curation, Investigation, Software.
RC: Conceptualization, Resources, Supervision, Validation, Writing – review & editing.
RS: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Acknowledgements

ZH, AR and RS acknowledge funding by the Swiss National Science Foundation (project MUTAMUR; no. 213976). We thank Riccardo Corazza for technical support, and Ignacio Pérez Prat for advice on data licensing.

References

Jachen Curdin Arquint. 1982. *Die viersprachige Schweiz*. Benziger.

Bernard Cathomas. 2024. *Ein Weg zur Einheit in der Vielfalt: Plädoyer für Rumantsch Grischun*. Somedia Buchverlag.

Rico M. Cathomas. 2005. Schule und Zweisprachigkeit: Immersiver Unterricht. *Internationaler Forschungsstand und eine empirische Studie am Beispiel des rätoromanisch-deutschen Schulmodells in der Schweiz*.

Eva Caviezel. 1993. Geschichte von Verschriftung, Normierung und Standardisierung des Surselvischen. *Societad Retorumantscha*.

Renata Coray. 2010. Rumantsch Grischun: Sprach-und Machtpolitik in Graubünden. *Annalas da la Societad Retorumantscha*, 123:147–165.

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.

Erwin Diekman. 1991. Probleme und Aspekte von Kodifizierungsbemühungen des Bündnerromanischen und Bericht über eine Umfrage zur Rezeption und Akzeptanz des Rumantsch Grischun als gesamtbündnerromanischer Schriftsprache. *W. Dahmen, O. Gsell, G. Holtus, J. Kramer, M. Metzeltin, & O. Winkelmann (Eds.), Zum Stand der Kodifizierung romanischer Kleinsprachen: Romanistisches Kolloquium*, V:69–104.

Eyal Liron Dolev. 2023. Does mBERT understand Romansh? evaluating word embeddings using word alignment. In *Proceedings of the 8th edition of the Swiss Text Analytics Conference*, pages 41–53, Neuchatel, Switzerland. Association for Computational Linguistics.

Barbla Etter. 2018. Widersprüche zwischen gesetzlich festgelegten Sprachgrenzen und der Sprachpraxis. *bulletin vals-asla numéro 108*, pages 35–54.

Jean-Jacques Furer. 2005. *Die aktuelle Lage des Romanischen*. Office fédéral de la statistique.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2022. The Flores-101 evaluation benchmark for low-resource and multilingual machine translation. *Transactions of the Association for Computational Linguistics*, 10:522–538.

Juri Grosjean and Jannis Vamvas. 2024. Fine-tuning the SwissBERT encoder model for embedding sentences and documents. In *Proceedings of the 9th edition of the Swiss Text Analytics Conference*, pages 41–49, Chur, Switzerland. Association for Computational Linguistics.

Manfred Gross. 1999. Rumantsch Grischun: Planification de la normalisation. *Bulletin suisse de linguistique appliquée*, (69):95–105.

Manfred Gross. 2017. Romansh: The Romansh language in education in Switzerland. *Mercator European Research Centre on Multilingualism and Language Learning*.

Matthias Grünert. 2024. Rätoromanisch. *Sprachenräume der Schweiz, Band 1: Sprachen*, pages 156–184.

Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Marzena Karpinska, Philipp Koehn, Benjamin Marie, Christof Monz, Kenton Murray, Masaaki Nagata, Martin Popel, Maja Popović, and 3 others. 2024. Findings of the WMT24 general machine translation shared task: The LLM era is here but MT is not solved yet. In *Proceedings of the Ninth Conference on Machine Translation*, pages 1–46, Miami, Florida, USA. Association for Computational Linguistics.

Julia Kreutzer, Isaac Caswell, Lisa Wang, Ahsan Wahab, Daan Van Esch, Nasanbayar Ulzii-Orshikh, Allahsera Tapo, Nishant Subramani, Artem Sokolov, Claytone Sikasote, and 1 others. 2022. Quality at a glance: An audit of web-crawled multilingual datasets. *Transactions of the Association for Computational Linguistics*, 10:50–72.

Jinhyuk Lee, Feiyang Chen, Sahil Dua, Daniel Cer, Madhuri Shanbhogue, Iftekhar Naim, Gustavo Hernández Ábrego, Zhe Li, Kaifeng Chen, Henrique Schechter Vera, and 1 others. 2025. Gemini embedding: Generalizable embeddings from gemini. *arXiv preprint arXiv:2503.07891*.

Ricarda Liver. 2010. *Rätoromanisch: eine Einführung in das Bündnerromanische*, second edition. Narr.

Fiona Müller and Maik Roth. 2019. *Sprachliche Praktiken in der Schweiz. Erste Ergebnisse der Erhebung zur Sprache, Religion und Kultur*. Bundesamt für Statistik.

Mathias Müller, Annette Rios, and Rico Sennrich. 2020. Domain robustness in neural machine translation. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 151–164, Virtual. Association for Machine Translation in the Americas.

Nathan Ng, Kyra Yee, Alexei Baevski, Myle Ott, Michael Auli, and Sergey Edunov. 2019. Facebook FAIR's WMT19 news translation task submission.

In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 314–319, Florence, Italy. Association for Computational Linguistics.

OpenAI, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, and 400 others. 2024. GPT-4o system card. *Preprint*, arXiv:2410.21276.

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.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Heinrich Schmid. 1989. Richtlinien für die Gestaltung einer gesamtbündnerromanischen Schriftsprache Rumantsch Grischun. *Annalas Da La Societad Retorumantscha*, 102:43–49.

Holger Schwenk, Vishrav Chaudhary, Shuo Sun, Hongyu Gong, and Francisco Guzmán. 2021. WikiMatrix: Mining 135M parallel sentences in 1620 language pairs from Wikipedia. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1351–1361, Online. Association for Computational Linguistics.

Edoardo Signoroni and Pavel Rychlý. 2023. Evaluating sentence alignment methods in a low-resource setting: An English-YorùBá study case. In *Proceedings of the Sixth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2023)*, pages 123–129, Dubrovnik, Croatia. Association for Computational Linguistics.

Arnold Spescha. 1989. *Grammatica sursilvana*. Casa editura per mieds d'instrucziun.

Brian Thompson and Philipp Koehn. 2019. Vecalign: Improved sentence alignment in linear time and space. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1342–1348, Hong Kong, China. Association for Computational Linguistics.

Jannis Vamvas, Johannes Graën, and Rico Sennrich. 2023. SwissBERT: The multilingual language model for Switzerland. In *Proceedings of the 8th edition of the Swiss Text Analytics Conference*, pages 54–69, Neuchatel, Switzerland. Association for Computational Linguistics.

Mike Zhang and Antonio Toral. 2019. The effect of translationese in machine translation test sets. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 73–81, Florence, Italy. Association for Computational Linguistics.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. 2025. Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models. *arXiv preprint arXiv:2506.05176*.

A Details on Choice of Embedding Model

We carry out a small-scale alignment experiment with our manually-aligned validation set to determine which model should be used for aligning *Mediomatix*. Specifically, for each of the idiom pairs, we embed the validation set and align a segment in the source idiom to the segment in the target idiom with the highest cosine similarity. We score the alignment in terms of the average proportion of correct one-to-one alignments across all idiom pairs.

We experiment with the following models for embedding the segments: Qwen3-Embedding-0.6B (Zhang et al., 2025), sentence-swissBERT (Grosjean and Vamvas, 2024), OpenAI's text-embedding-3-large[10], Google's gemini-embedding-exp-03-07 (Lee et al., 2025), Voyage AI's voyage-3-large[11], and Cohere's embed-v4.0[12]. Besides sentence-swissBERT—for which we explicitly run inference with the Romansh language adapter—we do not find information stating explicitly that Romansh is in the models' pretraining data. As in previous work on compiling parallel corpora for low-resource languages, we expect cross-lingual transfer in the multilingual models to support a reasonable embedding for Romansh data (Thompson and Koehn, 2019; Schwenk et al., 2021).

As noted in Section 3.1, we extract the plain text from the textbooks. In the plain text, we retain only the "<strong>" markup, as it sometimes provides insight about the meaning of a segment (i.e., in a question and answer with multiple choices, it may distinguish the correct choice). However, we also experiment with embedding the full HTML markup for each segment, as the structural similarities between segments in different textbooks encoded in the HTML may benefit alignment in case the Romansh text embeddings are not of sufficient quality on their own. In addition to embedding each segment's extracted text and full HTML, we experiment with concatenating the HTML and plain text embeddings. Results are in Table 4.

Model	Text	HTML	Concat
cohere-v4	96.2	94.1	95.7
gemini-embedding	96.2	95.6	96.3
openai-v3	93.8	90.1	93.5
qwen3-Embedding-0.6B	94.9	94.4	95.6
sentence-swissbert	73.7	70.3	77.3
voyage-v3	95.0	94.4	96.1

Table 4: Average proportion of correct alignments across all idiom pairs using different embedding models and text embeddings, HTML embeddings, and their concatenation ('Concat').

[10] https://platform.openai.com/docs/guides/embeddings/embedding-models

[11] https://blog.voyageai.com/2025/01/07/voyage-3-large/

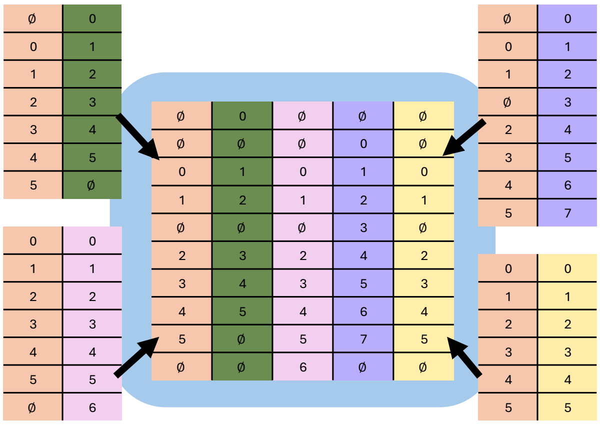
[12] https://docs.cohere.com/docs/cohere-embed

B Validation Set Statistics

	Segments	Tokens	Single Rows	Deletions	Many Rows
Sursilvan	155	1855	139	4	8
Sutsilvan	155	1846	136	6	9
Surmiran	151	1952	143	4	4
Puter	156	1849	140	4	7
Vallader	155	1842	139	4	8

Table 5: Descriptive statistics for the manually aligned multi-parallel validation set. “Single Rows” refers to alignment rows with just one segment, while “Many Rows” refers to alignment rows with more than one segment. “Deletions” refers to the number of rows for which the idiom has no parallel segment.

C Multi-parallel Alignment via a Pivot



∅	0
0	1
1	2
2	3
3	4
4	5
5	∅

0	0
1	1
2	2
3	3
4	4
5	5
∅	6

∅	0	∅	∅	∅
∅	∅	∅	0	∅
0	1	0	1	0
1	2	1	2	1
∅	∅	∅	3	∅
2	3	2	4	2
3	4	3	5	3
4	5	4	6	4
5	∅	5	7	5
∅	∅	6	∅	∅

∅	0
0	1
1	2
∅	3
2	4
3	5
4	6
5	7

0	0
1	1
2	2
3	3
4	4
5	5

Figure 2: Depiction of multi-parallel alignment via a pivot idiom, as described in Section 3.2. The four two-column tables represent bilingual alignments with a given pivot idiom (highlighted in orange). The five-column table shows the result of the full outer join of the bilingual alignments on the pivot idiom, which we take as the multi-parallel alignment for this pivot idiom.

D Validation Results for Individual Pivots

	Sursilvan	**Sutsilvan**	**Surmiran**	**Puter**	**Vallader**	**Consensus**
cohere-v4 (text)	94.8/99.2/96.9	93.3/98.4/95.7	92.3/97.4/94.7	95.1/99.6/97.2	94.4/99.2/96.7	97.2/94.1/95.4
gemini-embedding (concat)	93.9/98.6/96.1	92.8/98.0/95.2	91.2/96.5/93.7	95.4/99.8/97.5	94.4/99.2/96.7	98.1/92.1/94.8
voyage-v3 (concat)	94.8/99.2/96.9	93.7/98.8/96.1	93.2/98.2/95.5	95.2/99.6/97.2	94.4/99.2/96.7	97.3/95.6/96.3

Table 6: Average strict precision, recall, and F1 values for the validation set pivot alignments using the best three embedding models and input types from the greedy alignment experiment. The column represents the pivot idiom used. Underlined scores: Validation set scores for the final configuration used to align *Mediomatix*. We calculate values with the implementation provided with Vecalign, using the *strict* evaluation setup: https://github.com/thompsonb/vecalign

E *Mediomatix* Split Statistics

The test and validation splits make up a relatively large portion of the corpus. The primary motivation for splitting along grade levels rather than relegating a certain proportion of segments to each split was to reduce the likelihood of content overlap in the splits. A large validation and test set may also have practical benefits given that in the future, applications of the dataset may include QA or grammar exercise

Idiom	**Book Volumes**	**Aligned Segments**	**Aligned Tokens**
Sursilvan	16 / 10 / 8 / 33	12,071 / 7,329 / 4,764 / 25,708	103,416 / 80,815 / 57,351 / 271,454
Sutsilvan	16 / 10 / 8 / 33	11,621 / 7,355 / 4,753 / 25,408	100,788 / 83,126 / 58,763 / 271,173
Surmiran	11 / 8 / 8 / 0	6,033 / 3,906 / 4,362 / 0	54,273 / 38,888 / 52,821 / 0
Puter	16 / 10 / 8 / 31	11,684 / 7,355 / 4,660 / 23,486	103,991 / 81,698 / 57,037 / 252,124
Vallader	16 / 10 / 8 / 31	11,654 / 7,391 / 4,680 / 23,672	103,039 / 82,583 / 57,214 / 252,568

Table 7: Counts for the train/validation/test/no-rm-surmiran splits of the aligned *Mediomatix* corpus.

extraction. Large validation and test sets are needed for such tasks to ensure there is sufficient quantity of the exercises in the data used to test systems.

We also note that in Tables 1 and 7, the total number of book volumes in the aligned *Mediomatix* includes fewer volumes relative to the total number available in the schoolbook series (Appendix J). While manually aligning chapter titles between the books, we observed that some volumes contained no parallel content with the other idioms. In these cases, we dropped those book volumes and only automatically aligned sentences in volumes for idioms that did have largely comparable content.

F Qualitative Analysis of *Mediomatix*

	Example 1	**Example 2**	**Example 3**	**Example 4**
Sursilvan	ils 24 da fenadur tochen	Tgei munta la colur?	**La polizia ei stada tier els a casa.**	Jeu mon tuttina a prender penetenzia, schegie che jeu hai fatg ina massa puccaus.
Sutsilvan	igls 24 da fanadur antocen	Tge mùnta la calur?	**La polizeia e stada a controlar igl pass.**	Jou vont tutegna savens an la stizùn digl mazler, schagea ca jou magl bugent tgarn.
Surmiran	**igls 24 da fanadour**	Tge monta la calour?	**Las pulicistas òn controllo igl trafic.**	Ia magl tschigulatta, perchegl tg'ia va betg gugent ella.
Puter	ils 24 lügl fin	**Che at disch la culur da tias beschas?**	La polizia ho controllo il trafic.	Eau vegn listess a kino, eir scha muossan hoz bgers films.
Vallader	ils 24 lügl fin	**Che at disch la culur da tias beschas?**	La polizia ha controllà il trafic.	Eu vegn listess a kino, eir schi muossan hoz blers films.
Evaluator Comment	"*Less Information*"	"*More Information*"	"*Different examples of what the police are doing*"	"*Alignment Error*"

Table 8: Example multi-parallel alignments (columns) from the *Mediomatix* corpus. Segments in **bold** represent segments that were marked as noisy by the evaluator, while the segment in red text was marked as misaligned by the evaluator.

During their evaluation of the 100 randomly selected multi-parallel rows from *Mediomatix*, the annotator made several remarks. As shown in the second and third column of Table 8, they noted that when there was noise in a multi-parallel alignment, the Puter and Vallader segments were usually still parallel. They also noted that in some rows, when segments were not perfectly semantically parallel, they often still demonstrated the same spelling or grammar rule across idioms (i.e., conjugating different verbs for the conditional mood). Feedback such as that in the first two columns of Table 8 demonstrates a type of alignment noise in which one or more segments in a row contained slightly more or less information than the other segments.

G LLM Few-shot Prompt Example

~~~~
[{"role": "user",
  "content": "Translate the following segment surrounded in triple backticks
      into Vallader. The Sursilvan segment: \n```• Per tgei va ei en tiu
      cudisch? Resumescha quei che ti has legiu en treis
      construcziuns.```\n"},

 {"role": "assistant",
  "content": "```• Da che tratta teis cudesch? Resüma quai cha tü hast let
      in trais frasas.```"},

 {"role": "user",
  "content": "Translate the following segment surrounded in triple backticks
      into Vallader. The Sursilvan segment: \n```4. Jeu hai tschun sutbiadis
      che stattan el Grischun.```\n"},

 {"role": "assistant",
  "content": "```4. Eu n'ha tschinch bisabiadis chi stan in Grischun.```"},

 {"role": "user",
  "content": "Translate the following segment surrounded in triple backticks
      into Vallader. The Sursilvan segment: \n```p.ex. alla staziun, alla
      plazza aviatica, contact cun passagiers jasters```\n"},

 {"role": "assistant",
  "content": "```p.ex. a la staziun, a la plazza aviatica, contact cun
      passagers esters```"},

 {"role": "user",
  "content": "Translate the following segment surrounded in triple backticks
      into Vallader. The Sursilvan segment: \n```4. a) Tgeinina ei la
      differenza denter ils bustabs digl alfabet ed ils suns
      specials?```\n"}]
~~~~

H Detailed MT Results

H.1 NLLB-200-1.3B (fine-tuned)

Source → Target	Sursilvan	Sutsilvan	Surmiran	Puter	Vallader
Sursilvan	–	59.7	44.3	48.0	48.0
Sutsilvan	61.3	–	43.7	47.4	47.9
Surmiran	48.2	46.4	–	47.7	48.0
Puter	46.7	45.7	44.9	–	75.0
Vallader	47.1	46.1	45.4	74.1	–

Table 9: BLEU scores for machine translation between each pair of Romansh idioms in the *Mediomatix* corpus. Each cell shows the BLEU score for translating from the source idiom (row) to the target idiom (column).

H.2 GPT-4o (few-shot)

Source → Target	Sursilvan	Sutsilvan	Surmiran	Puter	Vallader
Sursilvan	–	24.6	25.4	33.1	40.7
Sutsilvan	59.9	–	25.7	29.3	37.0
Surmiran	48.9	23.2	–	33.9	41.1
Puter	43.9	19.9	25.5	–	65.7
Vallader	44.4	19.5	25.5	51.5	–

Table 10: BLEU scores for machine translation between each pair of Romansh idioms in the *Mediomatix* corpus. Each cell shows the BLEU score for translating from the source idiom (row) to the target idiom (column).

H.3 GPT-4o-mini (few-shot)

Source → Target	Sursilvan	Sutsilvan	Surmiran	Puter	Vallader
Sursilvan	–	29.9	27.1	25.9	27.1
Sutsilvan	38.0	–	28.0	25.4	26.7
Surmiran	32.3	27.4	–	26.2	28.7
Puter	28.1	24.0	23.9	–	57.0
Vallader	28.9	23.6	25.0	54.1	–

Table 11: BLEU scores for machine translation between each pair of Romansh idioms in the *Mediomatix* corpus. Each cell shows the BLEU score for translating from the source idiom (row) to the target idiom (column).

H.4 GPT-4o-mini (fine-tuned)

Source → Target	Sursilvan	Sutsilvan	Surmiran	Puter	Vallader
Sursilvan	–	39.9	35.3	32.9	34.2
Sutsilvan	47.9	–	34.5	33.5	34.1
Surmiran	38.1	31.0	–	32.8	34.9
Puter	35.4	29.1	33.1	–	67.1
Vallader	36.1	28.6	32.9	66.5	–

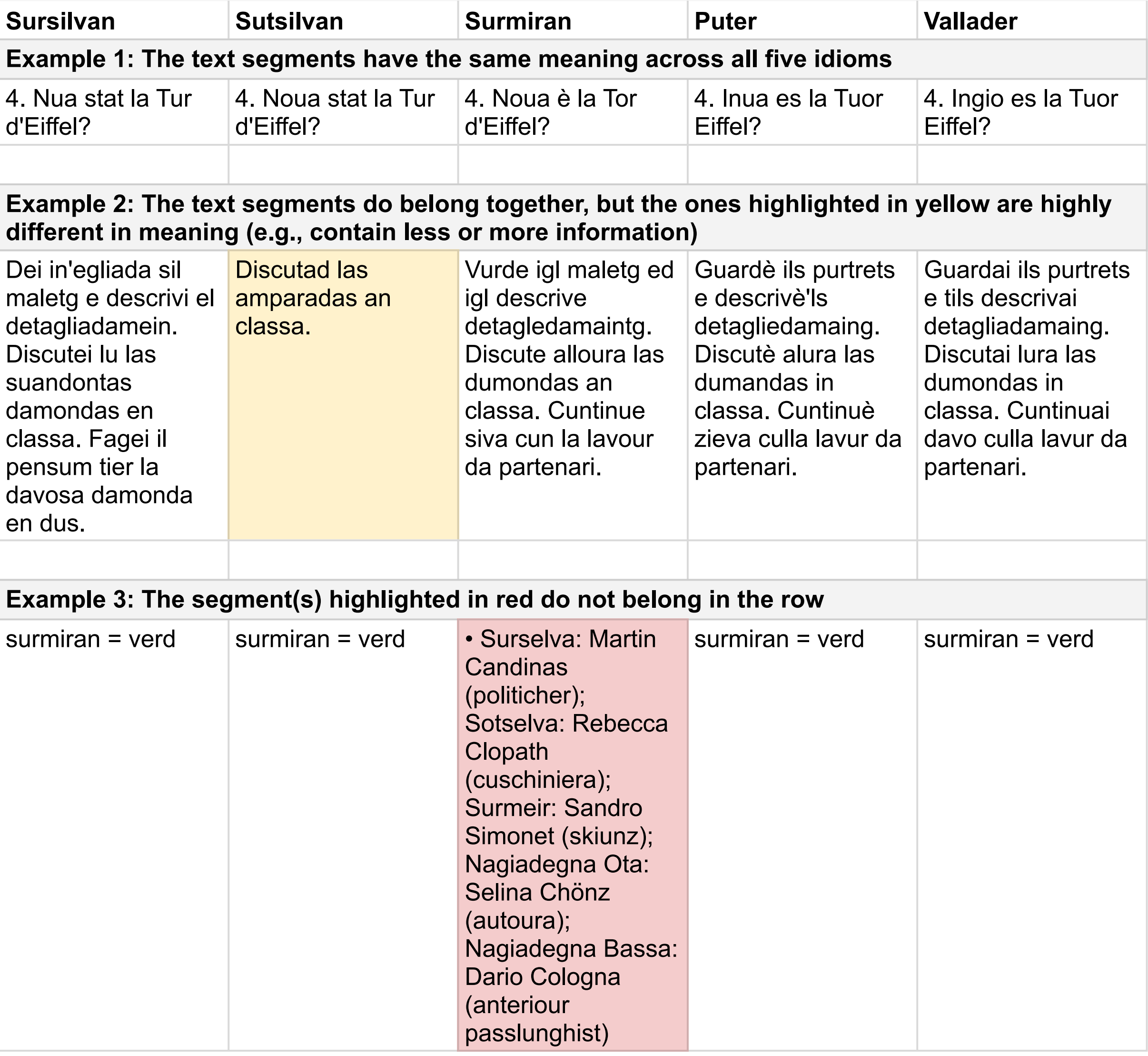
Table 12: BLEU scores for machine translation between each pair of Romansh idioms in the *Mediomatix* corpus. Each cell shows the BLEU score for translating from the source idiom (row) to the target idiom (column).

I Evaluator Instructions

You don't have to edit or proofread the examples. We'd just like to know whether our alignment algorithm worked, i.e., whether the text segments of the different idioms indeed belong together.

- If the text segments have the same meaning across all five idioms, which is the expected case, do nothing. See Example 1 in the Google Sheet.
- If the text segments do belong together, but there is an outlier that is highly different in meaning (e.g., contains different information), color it yellow. See Example 2 in the sheet.
- If the row has an outlier that clearly does not belong with the others, color it red. See Example 3 in the sheet.

Some cells will be empty if we did not find a matching segment for an idiom. This is okay and you don't have to mark the empty cells as errors.



Sursilvan	**Sutsilvan**	**Surmiran**	**Puter**	**Vallader**
Example 1: The text segments have the same meaning across all five idioms				
4. Nua stat la Tur d'Eiffel?	4. Noua stat la Tur d'Eiffel?	4. Noua è la Tor d'Eiffel?	4. Inua es la Tuor Eiffel?	4. Ingio es la Tuor Eiffel?
Example 2: The text segments do belong together, but the ones highlighted in yellow are highly different in meaning (e.g., contain less or more information)				
Dei in'egliada sil maletg e descrivi el detagliadamein. Discutei lu las suandontas damondas en classa. Fagei il pensum tier la davosa damonda en dus.	Discutad las amparadas an classa.	Vurde igl maletg ed igl descrive detagledamaintg. Discute alloura las dumondas an classa. Cuntinue siva cun la lavour da partenari.	Guardè ils purtrets e descrivè'ls detagliedamaing. Discutè alura las dumandas in classa. Cuntinuè zieva culla lavur da partenari.	Guardai ils purtrets e tils descrivai detagliadamaing. Discutai lura las dumondas in classa. Continuai davo culla lavur da partenari.
Example 3: The segment(s) highlighted in red do not belong in the row				
surmiran = verd	surmiran = verd	• Surselva: Martin Candinas (politicher); Sotselva: Rebecca Clopath (cuschiniera); Surmeir: Sandro Simonet (skiunz); Nagiadegna Ota: Selina Chönz (autoura); Nagiadegna Bassa: Dario Cologna (anteriour passlunghist)	surmiran = verd	surmiran = verd

J List of Schoolbooks

Bibliographic data of the schoolbooks included in the corpus, based on *Bündner Bibliografie*.
Each item in the list comprises 4 workbook volumes and 4 separate teacher's commentaries, except schoolbooks for the 4th and 6th grades, which have 5 volumes instead of 4.

Sursilvan

ISBN: 978-3-03847-012-0
Year: 2018
Title: Mediomatix, 2. classa, lungatg: sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006896790206696

ISBN: 978-3-03847-016-8
Year: 2019
Title: Mediomatix, 3. classa, lungatg sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007014990206696

ISBN: 978-3-03847-020-5
Year: 2020
Title: Mediomatix, 4. classa, lungatg: sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007140230206696

ISBN: 978-3-03847-024-3
Year: 2021
Title: Mediomatix, 5. classa, lungatg: sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997476532206696

ISBN: 978-3-03847-028-1
Year: 2021
Title: Mediomatix, 6. classa, lungatg: sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997476532306696

ISBN: 978-3-03847-032-8
Year: 2020
Title: Mediomatix, 1. classa scalem secundar 1, lungatg: sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007112700206696

ISBN: 978-3-03847-036-6
Year: 2019
Title: Mediomatix, 2. classa scalem secundar 1, lungatg: sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007015660206696

ISBN: 978-3-03847-040-3
Year: 2018
Title: Mediomatix, 3. classa scalem secundar 1, lungatg: sursilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006897030206696

Sutsilvan

ISBN: 978-3-03847-013-7
Year: 2018
Title: Mediomatix, 2. classa, lungatg: sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006896810206696

ISBN: 978-3-03847-017-5
Year: 2019
Title: Mediomatix, 3. classa, lungatg sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007015060206696

ISBN: 978-3-03847-021-2
Year: 2020
Title: Mediomatix, 4. classa, lungatg: sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007140200206696

ISBN: 978-3-03847-025-0
Year: 2021
Title: Mediomatix, 5. classa, lungatg: sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997476532006696

ISBN: 978-3-03847-029-8
Year: 2021
Title: Mediomatix, 6. classa, lungatg: sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997476531806696

ISBN: 978-3-03847-033-5
Year: 2020
Title: Mediomatix, 1. classa scalem secundar 1, lungatg: sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007112660206696

ISBN: 978-3-03847-037-3
Year: 2019
Title: Mediomatix, 2. classa scalem secundar 1, lungatg: sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007015630206696

ISBN: 978-3-03847-041-0
Year: 2018
Title: Mediomatix, 3. classa scalem secundar 1, lungatg: sutsilvan.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006897020206696

Surmiran

ISBN: 978-3-03847-122-6
Year: 2022
Title: Mediomatix. 2, lungatg: surmiran.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997566920606696

ISBN: 978-3-03847-123-3
Year: 2023
Title: Mediomatix. 3. classa, lungatg: surmiran.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997922021006696

ISBN: 978-3-03847-124-0
Year: 2024
Title: Mediomatix. 4. classa, lungatg: surmiran.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma998005220706696

ISBN: 978-3-03847-125-7
Year: 2025
Title: Mediomatix. 5. classa, lungatg: surmiran.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma998115845006696

Puter

ISBN: 978-3-03847-014-4
Year: 2018
Title: Mediomatix, 2. classa, lungatg: puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006896830206696

ISBN: 978-3-03847-018-2
Year: 2019
Title: Mediomatix, 3. classa, lingua puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007015160206696

ISBN: 978-3-03847-022-9
Year: 2020
Title: Mediomatix, 4. classa, lingua: puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007140210206696

ISBN: 978-3-03847-026-7
Year: 2021
Title: Mediomatix, 5. classa, lingua: puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997476531706696

ISBN: 978-3-03847-030-4
Year: 2021
Title: Mediomatix. 6. classa lingua: puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997476531606696

ISBN: 978-3-03847-034-2
Year: 2020
Title: Mediomatix, 1. classa s-chelin secundar 1, lingua: puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007112690206696

ISBN: 978-3-03847-038-0
Year: 2019
Title: Mediomatix, 2. classa s-chelin secundar 1, lungatg: puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007015590206696

ISBN: 978-3-03847-042-7
Year: 2018
Title: Mediomatix, 3. classa s-chelin secundar 1, lungatg: puter.
Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006896870206696

Vallader

ISBN: 978-3-03847-015-1

Year: 2018

Title: Mediomatix, 2. classa, lungatg: vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006896860206696

ISBN: 978-3-03847-019-9

Year: 2019

Title: Mediomatix, 3. classa, lingua vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007014880206696

ISBN: 978-3-03847-023-6

Year: 2020

Title: Mediomatix, 4. classa, lingua: vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007140190206696

ISBN: 978-3-03847-027-4

Year: 2021

Title: Mediomatix, 5. classa, lingua: vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997481737406696

ISBN: 978-3-03847-031-1

Year: 2021

Title: Mediomatix, 6. classa, lingua: vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma997481737306696

ISBN: 978-3-03847-035-9

Year: 2020

Title: Mediomatix, 1. classa s-chalin secundar 1, lingua: vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007112650206696

ISBN: 978-3-03847-039-7

Year: 2019

Title: Mediomatix, 2. classa s-chalin secundar 1, lingua: vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990007015710206696

ISBN: 978-3-03847-043-4

Year: 2018

Title: Mediomatix, 3. classa s-chalin secundar 1, lungatg: vallader.

Bibliographic record: https://www.opac.gr.ch/permalink/41BGR_INST/44cnm/alma990006897040206696