Training an NMT system for legal texts of a low-resource language variety (South Tyrolean German – Italian)

Antoni Oliver and **Sergi Álvarez**
Universitat Oberta de Catalunya
{aoliverg,salvarezvid}
@uoc.edu

Egon W. Stemle and **Elena Chiocchetti**
Eurac Research
{egon.stemle,elena.chiocchetti}
@eurac.edu

Abstract

This paper illustrates the process of training and evaluating NMT systems for a language pair that includes a low-resource language variety. A parallel corpus of legal texts for Italian and South Tyrolean German has been compiled, with South Tyrolean German being the low-resourced language variety. As the size of the compiled corpus is insufficient for the training, we have combined the corpus with several parallel corpora using data weighting at sentence level. We then performed an evaluation of each combination and of two popular commercial systems.

1 Introduction

Neural machine translation (NMT) has shown outstanding performance and translation quality compared to previous models (Bentivogli et al., 2016). However, there is a translation quality gap to fill for low-resource languages (Aranberri and Iñurrieta, 2024; Goyle et al., 2023; Ranathunga et al., 2023) due to the significant amount of parallel data that is required to learn useful mappings between languages (Lakew et al., 2018). Besides, the legal domain poses an increased challenge for NMT (Killman, 2023) given the intricate nature of legal language, the necessity to use precise terminology, and the negative consequences of misunderstanding the legal intent (Quinci and Pontrandolfo, 2023).

Our work deals both with the legal domain and with a low-resource language variety of a big European language (German), as we translated decrees in the language combination Italian – South Tyrolean German. It shows some of the challenges of NMT in the legal domain, focusing on a low-resource variety. This low-resource variety context is not unique but shared by many language communities in relation to the legal and administrative domain; in Europe, for example, by the German-speaking community in Belgium, the Swedish-speaking community in Finland, the Italian-speaking community in Croatia, the Danish-speaking community in Germany and many more.

After having trained our NMT systems with the LEXB (Contarino, 2021) corpus by Eurac Research and having cleaned and curated the data afterwards, we achieved an increase of, depending on the direction of translation, approximately five to eight full points in BLEU (Post, 2018). This shows the importance of training domain-specific NMT systems with high-quality data, especially for low-resource languages.

1.1 Linguistic situation in South Tyrol

South Tyrolean German is the standard variety of German (Ammon et al., 2016) used in Northern Italy in the Autonomous Province of Bolzano (South Tyrol). German is an officially recognized minority language in South Tyrol. The public administration offices are legally bound to use German next to Italian when dealing with the citizens[1] (Presidential Decree No. 670/1972, Art. 99). All administrative documents, local legislation and material aimed at the general public (e.g. websites of local public institutions) must be available in

[1]The small Ladin-speaking minority (20,000 speakers) has also been granted extensive language rights. However, these are generally limited to the valleys of Gherdëina and Badia.

Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1), pages 573–579
June 24-27, 2024 ©2024 European Association for Machine Translation

Italian and German.

This officially multilingual institutional regime is implemented by translating texts from German into Italian or vice versa, increasingly using machine translation (De Camillis, 2021). Even though the majority (69%) of the South Tyrolean population is German-speaking (ASTAT – Provincial Statistics Institute, 2021) and today many legal texts are drafted in the minority language, only the Italian version of legal texts is legally binding in case of diverging interpretations (Presidential Decree No. 670/1972, Art. 99). This implies that it is a translated text—either translated or post-edited by a human—that often becomes the legally valid text in South Tyrol.

1.2 Challenges in legal translation

The consequences of mistakes in legal translation may be serious (Mattila, 2018) and include financial loss, legal disputes, infringement of basic human rights (e.g. bad interpreting in a criminal court case). Legal language is therefore considered particularly difficult to translate (Killman, 2023; Mattila, 2018). This is partly due to some specific characteristics of legal language, among others (Gualdo and Telve, 2021; Mattila, 2018):

- specific syntactic features and generally long and complex sentences;
- closeness to general language, with general language words often taking on a specific meaning in the legal context (e.g. 'trust');
- terminology that is system-bound and therefore may vary even across legal systems using the same language (e.g. 'antitrust law' in the US vs 'competition law' in the UK) and may additionally vary in meaning—and translation—also across legal subdomains (e.g. in US banking law, 'withdrawal' means the removal of money from a bank but in US criminal law it refers to a person separating themselves from criminal activity[2]; the first term can be translated with *prelievo*, the second with *dissociazione* in Italian);
- use of abbreviations, acronyms and initialisms;
- formulaic legal phraseology that should not be translated literally.

[2]https://thelawdictionary.org/withdrawal/

All these features are present in the Italian and South Tyrolean German legal languages and pose notable challenges to NMT systems. Chromá (2008) stressed the central role of terminology in legal translation by calculating that between 20% and 29% of legal texts consist of terminology.

1.3 South Tyrolean German and translation

South Tyrolean German has syntactic, grammatical and lexical features that generally characterize it as a Southern German variety and are often shared with the Austrian and/or Swiss standard varieties. Examples are the choice of the auxiliary to form the past tense of some verbs and the use of linking elements within compounds. However, its specific terminology in the domain of law and food as well as a significant influence of Italian clearly distinguishes it from the neighbouring varieties (Ammon et al., 2016). Heiss and Soffritti (2018) and Wiesmann (2019) have shown that terminology is also one of the major machine translation issues in the language combination Italian – South Tyrolean German. A more in-depth error annotation by De Camillis et al. (2023) found that mistranslations and bilingual terminology errors were the most represented error categories when machine-translating South Tyrolean legal texts.

Mistranslations comprise several subcategories of mistakes where the source meaning has been incorrectly transferred to the target language. These include multi-word expressions that have conventional—often non-literal—equivalents like collocations and titles of laws, polysemous words that were disambiguated in the wrong way, occurrences of translations with semantically unrelated words and instances of errors in translating gender-sensitive language. The latter is a known bias of NMT systems (Savoldi et al., 2021). The local South Tyrolean legislation must be inclusive of all genders or at least inclusive of the male and female genders (Provincial Law No. 5/2010, Art. 8). This is achieved by using gender-neutral formulations or terms (e.g. *Lehrperson*, 'teaching person') and split forms mentioning both the male and female forms (e.g. *Lehrerinnen* und *Lehrer*, 'female and male teachers') in all language versions. Disrespecting this requirement by generally using only male terms as NMT systems often do (e.g. by translating a gender-neutral expression like *eine Lehrperson* with *un insegnante*, the male form of teacher in Italian) entails a breach of the

law and causes notable post-editing efforts.

Bilingual terminology errors relate to wrongly translated terms in general but also to improper use of terminology pertaining to other legal systems (e.g. *Land* translated with *stato*, 'state', rather than *provincia*, 'province', because the term refers to a federated state in Germany and Austria but to the Autonomous Province in South Tyrol). The consequences of mistranslating such terminology from German into Italian in South Tyrolean texts are a wrong attribution of competences to the state rather than to the provincial level of governance. Disrespecting the correct terminology does not necessarily make the NMT output impossible to understand. Many South Tyroleans would grasp the meaning of *Tarifvertrag* ('collective bargaining agreement' in Germany), even though the correct legal term in South Tyrol is *Kollektivvertrag*. It may also be relatively easy to amend for post-editors with good in-domain knowledge. However, using correct terminology is essential from a legal point of view. Incorrect terms create doubts as to which legal concept is referred to and which legal texts form the legal basis serve as reference. In addition, in a minority language situation, using inconsistent or incorrect legal terminology impairs legal certainty and discriminates against the members of the minority community, as the latter will face additional issues in understanding their legal texts compared to the members of the majority.

2 Experimental part

2.1 The LEXB Italian-German corpus by Eurac Research

Contarino (2021)'s LEXB corpus, a bilingual parallel corpus of Italian and South Tyrolean German, was slightly refined at Eurac Research. It features local and national legislation retrieved from the LexBrowser database[3], which gathers laws, decrees, resolutions, collective agreements and other national legal legislation of interest to South Tyrol. The corpus also contains a limited number of bilingual texts not published in the LexBrowser collection, namely 20 national laws and codes (Civil Code, Criminal Code) translated into German, mainly by the provincial Office for Language Issues. This original corpus data has been further cleaned for the current project using MTUOC-clean-parallel-corpus[4] and rescored with MTUOC-PCorpus-rescorer[5] (Oliver and Álvarez, 2023). The number of segments and tokens of the raw compiled corpus and the clean and rescored version is included in Table 1. As we can observe, the number of available parallel segments is inadequate for training an NMT system. For this reason, we have combined this corpus with several parallel corpora, as described in the following subsection.

Table 1: Size of the LEXB Italian-German parallel corpus by Eurac Research.

Corpus	Segments	tokens ita	tokens deu
raw	173,530	5,027,663	4,569,333
clean	164,291	4,882,422	4,438,953

Table 2: Size of the Italian-German parallel corpus downloaded from Opus Corpus.

Corpus	Segments
Multiparacrawl	30,337,479
EU rescored	4,936,565

2.2 Other Italian-German corpora used

Table 2 describes the corpora used and their respective sizes in unique segments and tokens. The EU corpus was obtained by concatenating and deduplicating the following corpora: DGT, ELRC-EMEA, EMEA, Europarl and JRC Acquis. The resulting corpus was cleaned and rescored. All the corpora included in this subsection were obtained from Opus Corpus[6] (Tiedemann, 2009).

2.3 Tools used to train the NMT systems

We used the following tools to train the NMT systems:

- To preprocess the corpora: MTUOC-corpus-preprocessing[7]. This tool allows to use, among other algorithms, sentencepiece[8] (Kudo and Richardson, 2018).
- Marian NMT[9] (Junczys-Dowmunt et al., 2018)

[3] http://lexbrowser.provinz.bz.it/
[4] https://github.com/mtuoc/MTUOC-clean-parallel-corpus
[5] https://github.com/mtuoc/MTUOC-PCorpus-rescorer
[6] https://opus.nlpl.eu/
[7] https://github.com/mtuoc/MTUOC-corpus-preprocessing
[8] https://github.com/google/sentencepiece
[9] https://marian-nmt.github.io/

Table 3: Evaluation results for the Italian-German NMT systems.

System	BLEU (μ ± 95% CI)	chrF2 (μ ± 95% CI)	TER (μ ± 95% CI)
Baseline: Multiparacrawl	37.8 (37.8 ± 1.6)	58.3 (58.3 ± 1.0)	57.0 (57.0 ± 1.7)
EU	29.6 (29.6 ± 1.2) (p = 0.0010)*	53.4 (53.4 ± 0.9) (p = 0.0010)*	60.8 (60.8 ± 1.2) (p = 0.0010)*
EURAC-EU	**52.5** (52.5 ± 2.0) (p = 0.0010)*	**68.6** (68.6 ± 1.1) (p = 0.0010)*	**42.0** (41.9 ± 1.9) (p = 0.0010)*
EURAC-EU-Multiparacrawl	47.9 (47.9 ± 1.9) (p = 0.0010)*	64.2 (64.2 ± 1.0) (p = 0.0010)*	45.5 (45.5 ± 1.5) (p = 0.0010)*
GoogleT	44.1 (44.1 ± 1.3) (p = 0.0010)*	65.6 (65.6 ± 0.7) (p = 0.0010)*	45.8 (45.8 ± 1.2) (p = 0.0010)*
DeepL	36.8 (36.8 ± 1.2) (p = 0.0849)	63.6 (63.6 ± 0.7) (p = 0.0010)*	51.3 (51.2 ± 1.1) (p = 0.0010)*

Table 4: Evaluation results for the German-Italian NMT systems.

System	BLEU (μ ± 95% CI)	chrF2 (μ ± 95% CI)	TER (μ ± 95% CI)
Baseline: Multiparacrawl	33.3 (33.3 ± 1.1)	56.9 (56.9 ± 0.9)	54.9 (54.9 ± 1.1)
EU	47.3 (47.3 ± 1.4) (p = 0.0010)*	67.9 (67.9 ± 1.0) (p = 0.0010)*	44.9 (44.9 ± 1.3) (p = 0.0010)*
EURAC-EU	**53.5** (53.5 ± 1.6) (p = 0.0010)*	**71.0** (71.0 ± 1.0) (p = 0.0010)*	**39.1** (39.1 ± 1.4) (p = 0.0010)*
EURAC-EU-Multiparacrawl	48.3 (48.3 ± 1.7) (p = 0.0010)*	66.1 (66.1 ± 1.1) (p = 0.0010)*	44.2 (44.2 ± 1.5) (p = 0.0010)*
GoogleT	43.5 (43.5 ± 1.1) (p = 0.0010)*	68.0 (68.0 ± 0.7) (p = 0.0010)*	44.0 (44.1 ± 1.1) (p = 0.0010)*
DeepL	47.6 (47.5 ± 1.3) (p = 0.0010)*	70.0 (70.0 ± 0.7) (p = 0.0010)*	42.1 (42.1 ± 1.1) (p = 0.0010)*

2.4 Training procedure

With the corpora described in Sections 2.1 and 2.2 and the tools described in Section 2.3, we trained the following systems in both directions (Italian-German and German-Italian):

- Multiparacrawl: these are the baseline systems trained using the Multiparacrawl corpus (see Section 2.2).
- EU: these systems were trained using the EU corpus (see Section 2.2).
- EURAC-EU: these systems were trained using the EURAC corpus (see Section 2.1) with a sentence weight of 1 and the EU corpus with a sentence weight of 0.5.
- EURAC-EU-Mutliparacrawl: these systems were trained using the EURAC corpus with a sentence weight of 1, the EU corpus with a sentence weight of 0.5 and the Multiparacrawl corpus with a sentence weight of 0.25.

All the corpora have been split into training, validation and evaluation parts. As corpora have been deduplicated, no common segments are present in these subsets. Validation and evaluation sets are formed by 5,000 segments each, and the rest of the segments are used in the training subset. For the EURAC-EU and EURAC-EU-Multiparacrawl corpora, the validation and evaluation subset segments are selected from the EURAC corpus.

All the training processes were performed on a computer with 2 GPUs NVIDIA RTX A 5000 with 24GB each, with the following parameters:

- Guided alignment using eflomal[10] (Östling and Tiedemann, 2016).
- Size of vocabularies: 32,000
- Valid metrics: cross-entropy and bleu-detok
- Patience: 10 on all metrics.
- Type of model: transformer
- Max length of training segments: 150 tokens.

2.5 Evaluation

To evaluate the trained systems, we have used 1,000 segments from the evaluation sets. The trained systems were evaluated along with two popular commercial NMT systems: Google Translate[11] and DeepL[12]. We accessed both commercial systems through their respective APIs using Python scripts.

For the evaluation, we used three automatic metrics implemented in Sacrebleu[13] (Post, 2018): BLEU, chrF2 and TER. The appendices present the signatures of the three metrics stating the exact configuration parameters as reported by Sacrebleu.

Tables 3 and 4 show the evaluation results for the Italian-German and German-Italian systems. In both cases, the baseline systems are trained using only the Multiparacrawl corpus. In the evaluation, a paired bootstrap resampling test with 1,000 resampling trials was performed using the `-paired-bs option` in Sacrebleu. In this way, each system is pairwise compared to the baseline system Multiparacrawl. Assuming a significance threshold of 0.05, the null hypothesis can be rejected for p-values < 0.05 (marked with "*" in the tables.)

For both language pairs, the best-performing system according to the used automatic metrics is the systems trained using the EURAC and the EU corpora. For the Italian-German pair, the system improves the baseline system by 14.7 BLEU points, Google Translate by 8.4 BLEU points and DeepL by 15.7 BLEU points. For the rest of the automatic metrics, this system also outperforms the baseline and the commercial systems. The same happens for the German-Italian language pair, where the EURAC-EU system improves the baseline, Google Translate and DeepL by 20.2, 10 and 5.9 BLEU points, respectively.

Table 5 shows the improvements achieved by the EURAC-EU systems compared with the two commercial systems, Google Translate and DeepL, along with the statistical significance test results for both Italian-German and German-Italian. Figures in the table show the increment of BLEU and chrF2, as well as the decrement of TER, as lower TER values indicate better quality. As we can see in the table, the EURAC-EU systems outperform the commercial systems for the two language pairs and for all the automatic metrics. All these results pass the statistical significance test.

3 Conclusions and future work

Low-resource language situations are challenging for NMT engines. We are working with a low-resource language variety of a major European language and with legal texts, which in itself is a low-resource situation and additionally requires a very particular language.

We have trained an NMT model for the legal domain with the language combination Italian – South Tyrolean German, a low-resource language variety. To this end, we have used and processed a relevant available corpus of legal texts. As this in-domain corpus is not big enough to train NMT systems, we have augmented this data with combinations of other corpora: a corpus created from several EU corpora and Multiparacrawl. The combinations are based on weighting at sentence level, giving higher weight to segments from the compiled in-domain corpus. As a baseline system, we have trained an NMT system using the Multiparacrawl corpus only.

Results show that the best system is the one trained with the in-domain corpus combined with the EU corpora, as it performs better than commercial products for these language combinations. An evaluation was carried out using three of the most frequent assessment metrics (BLEU, chrF2, TER). As positive as these results may seem, a qualitative breakdown of the results, with manual annotations along established criteria (De Camillis et al., 2023) to better understand the specifics of the particular circumstances, is still pending and is planned as the next step.

This paper shows that training tailored NMT systems can be a viable alternative to commercial systems in a low-resource scenario. Even with lim-

[10] https://github.com/robertostling/eflomal
[11] https://translate.google.com/
[12] https://www.deepl.com/en/translator
[13] https://github.com/mjpost/sacrebleu

Table 5: Improvements and statistical significance of the EURAC-EU system vs Google Translate and DeepL.

L.P.	System	BLEU (μ ± 95% CI)	chrF2 (μ ± 95% CI)	TER (μ ± 95% CI)
ita-deu	GoogleT	+8.4 (p = 0.0010)*	+3.0 (p = 0.0010)*	-3.8 (p = 0.0010)*
ita-deu	DeepL	+15.7 (p = 0.0010)*	+5.0 (p = 0.0010)*	-9.3 (p = 0.0010)*
deu-ita	GoogleT	+10.0 (p = 0.0010)*	+3.0 (p = 0.0010)*	-4.9 (p = 0.0010)*
deu-ita	DeepL	+5.9 (p = 0.0010)*	+1.0 (p = 0.0010)*	-3.0 (p = 0.0010)*

ited in-domain data, using data from a similar domain and data weighting techniques, the final system can outperform widely used commercial systems. In particular, low-resource varieties of bigger languages tend to be neglected in research and NMT development, even though the consequences of mistranslation may be serious. With its system-bound terminology and phraseology, the legal domain needs particular attention, as it is relevant for legal and translation professionals increasingly using NMT systems and the general public.

Finally, our results emphasize the importance of curated in-domain corpora to align the results of NMT models with those pertaining to situations with more data.

Acknowledgements

This work has been done in the framework of the research and technology transfer agreement between Eurac Research (Bolzano, Italy) and the Universitat Oberta de Catalunya (UOC, Catalonia, Spain).

Appendices:

Metric signatures

- BLEU: nrefs:1 | bs:1000 | seed:12345 |case:mixed | eff:no | tok:13a | smooth:exp|version:2.3.1
- chrF2: nrefs:1 | bs:1000 | seed:12345 | case:mixed | eff:yes | nc:6 | nw:0 | space:no | version:2.3.1
- TER: nrefs:1 | bs:1000 | seed:12345 | case:lc | tok:tercom | norm:no | punct:yes | asian:no | version:2.3.1

References

Ammon, Ulrich, Hans Bickel, and Alexandra N. Lenz, editors. 2016. *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen*. de Gruyter, Berlin, 2 edition.

Aranberri, Nora and Uxoa Iñurrieta. 2024. When minoritized languages encounter MT: perceptions and expectations of the Basque community. *Jostrans – The Journal of Specialised Translation*, (41):179–205.

ASTAT – Provincial Statistics Institute. 2021. *South Tyrol in Figures*. Provincial Statistics Institute, Bolzano/Bozen.

Bentivogli, Luisa, Arianna Bisazza, Mauro Cettolo, and Marcello Federico. 2016. Neural versus phrase-based machine translation quality: a case study. In Su, Jian, Kevin Duh, and Xavier Carreras, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 257–267, Austin, Texas, November. Association for Computational Linguistics.

Chromá, Marta. 2008. Translating Terminology in Arbitration Discourse. In Bhatia, Vijay K., Christopher N. Candlin, Jan Engberg, and Jane Lung, editors, *Legal Discourse across Cultures and Systems*, pages 309–328. Hong Kong University Press, Hong Kong.

Contarino, Antonio. 2021. *Neural Machine Translation Adaptation and Automatic Terminology Evaluation: A Case Study on Italian and South Tyrolean German Legal Texts*. Ph.D. thesis, Università di Bologna, Bologna, Italy.

De Camillis, Flavia, Egon Stemle, Elena Chiocchetti, and Francesco Fernicola. 2023. The MT@BZ corpus: machine translation & legal language. *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 171–180.

De Camillis, Flavia. 2021. *La traduzione non professionale nelle istituzioni pubbliche dei territori di lingua minoritaria: il caso di studio dell'amministrazione della Provincia autonoma di Bolzano*. Dissertation, Università di Bologna.

Goyle, Vakul, Parvathy Krishnaswamy, Kannan Girija Ravikumar, Utsa Chattopadhyay, and Kartikay Goyle. 2023. Neural machine Translation for low resource languages.

Gualdo, Riccardo and Stefano Telve. 2021. *Linguaggi specialistici dell'italiano*. Carocci, Roma.

Heiss, Christine and Marcello Soffritti. 2018. DeepL Traduttore e didattica della traduzione dall'italiano in tedesco. *inTRAlinea*, 20(1).

Junczys-Dowmunt, Marcin, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, and Alexandra Birch. 2018. Marian: Fast neural machine translation in C++. In *Proceedings of ACL 2018, System Demonstrations*, pages 116–121, Melbourne, Australia, July. Association for Computational Linguistics.

Killman, Jeffrey. 2023. Machine translation and legal terminology. Data-driven approaches to contextual accuracy. In Biel, Łucja and Hendrik J. Kockaert, editors, *Handbook of Terminology. Legal Terminology*, volume 3, pages 485–510. Benjamins, Amsterdam / Philadelphia.

Kudo, Taku and John Richardson. 2018. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71.

Lakew, Surafel M., Marcello Federico, Matteo Negri, and Marco Turchi. 2018. Multilingual Neural Machine Translation for Low-Resource Languages. *IJCoL. Italian Journal of Computational Linguistics*, 4(1):11–25, June. Number: 1 Publisher: Accademia University Press.

Mattila, Heikki E.S. 2018. Legal Language. In Humbley, John, Gerhard Budin, and Christer Laurén, editors, *Languages for Special Purposes: An International Handbook*, pages 113–150. De Gruyter Mouton, Berlin, Boston.

Oliver, Antoni and Sergi Álvarez. 2023. Filtering and rescoring the CCMatrix corpus for neural machine translation training. In Nurminen, Mary, Judith Brenner, Maarit Koponen, Sirkku Latomaa, Mikhail Mikhailov, Frederike Schierl, Tharindu Ranasinghe, Eva Vanmassenhove, Sergi Alvarez Vidal, Nora Aranberri, Mara Nunziatini, Carla Parra Escartín, Mikel Forcada, Maja Popovic, Carolina Scarton, and Helena Moniz, editors, *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 39–45, Tampere, Finland, June. European Association for Machine Translation.

Östling, Robert and Jörg Tiedemann. 2016. Efficient word alignment with Markov Chain Monte Carlo. *Prague Bulletin of Mathematical Linguistics*, 106:125–146, October.

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

Quinci, Carla and Gianluca Pontrandolfo. 2023. Testing neural machine translation against different levels of specialisation: An exploratory investigation across legal genres and languages. *trans-kom*, 16:174–209, July.

Ranathunga, Surangika, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. 2023. Neural Machine Translation for Low-Resource Languages: A Survey.

Savoldi, Beatrice, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2021. Gender Bias in Machine Translation. In Roark, Brian and Ani Nenkova, editors, *Transactions of the Association for Computational Linguistics*, volume 9, pages 845–874, Cambridge. Association for Computational Linguistics.

Tiedemann, Jörg. 2009. News from OPUS: A collection of multilingual parallel corpora with tools and interfaces. In Nicolov, N., K. Bontcheva, G. Angelova, and R. Mitkov, editors, *Recent Advances in Natural Language Processing*, volume V, pages 237–248. John Benjamins.

Wiesmann, Eva. 2019. Machine translation in the field of law: A study of the translation of Italian legal texts into German. *Comparative Legilinguistics. International Journal for Legal Communication*, 37:117–153.