

From Zero to Production: Baltic-Ukrainian Machine Translation Systems to Aid Refugees

Toms BERGMANIS, Mārcis PINNIS

Tilde, Vienības gatve 75A, Riga, Latvia, LV-1004

`{name.surname}@tilde.lv`

Abstract. In this paper, we examine the development and usage of six low-resource machine translation systems translating between the Ukrainian language and each of the official languages of the Baltic states. We developed these systems in reaction to the escalating Ukrainian refugee crisis caused by the Russian military aggression in Ukraine in the hope that they might be helpful for refugees and public administrations. Now, two months after MT systems were made public, we analyze their usage patterns and statistics. Our findings show that the Latvian-Ukrainian and Lithuanian-Ukrainian systems are integrated into the public services of Baltic states, leading to more than 127 million translated sentences for the Lithuanian-Ukrainian system. Motivated by these findings, we further enhance our MT systems by better Ukrainian toponym translation and publish an improved version of the Lithuanian-Ukrainian system.

Keywords: low resource machine translation, Ukrainian, Latvian, Lithuanian, Estonian

1 Introduction

On February 20, 2014, Russian Federation started military aggression against Ukraine (Cosgrove, 2020). Eight years later, on February 24, 2022, following a Russian military build-up on the Russia–Ukraine border, Russian aggression culminated in a full-scale invasion of Ukraine.¹ As of May 2022, more than 6.1 million refugees have fled Ukraine.² The majority of refugees have left Ukraine for one of the seven neighboring countries. Still, many seek shelter in other countries, including the Baltic states.³ The influx of Ukrainian refugees poses a new challenge for communication between individuals and governmental bodies in the Baltic states.

¹ Accessed May 16, 2022 <https://news.un.org/en/focus/ukraine>

² Accessed May 16, 2022 <https://data2.unhcr.org/en/situations/ukraine>

³ Accessed May 5, 2022 https://en.wikipedia.org/wiki/2022_Ukrainian_refugee_crisis#Other_European_countries

In this paper, we examine six low-resource machine translation (MT) systems translating between the Ukrainian language and each of the official languages of the Baltic states. Their development took place in the wake of the escalating Ukrainian refugee crisis shortly after the Russian invasion of Ukraine. Thus it was motivated by apprehension for the future rather than a clear vision of how they might be used. Now, after MT systems have been online for more than two months, we analyze their usage statistics and draw conclusions for what are the aspects of MT integration in the public services, which have led to more than 127 million translated sentences for the Lithuanian-Ukrainian system, while the Latvian-Ukrainian system has been used seemingly relatively little having translated only 138 thousand sentences.

2 Machine Translation Systems

Due to data scarcity for the language pairs involving Ukrainian and the languages of the Baltic states, we use two data augmentation methods – one that enables dynamic terminology integration and another that allows training MT models that are more robust to unknown tokens and rare words. For terminology integration, we prepare data with Target Lemma Annotations (TLA) (Bergmanis and Pinnis, 2021b), while for the robustness, we use synthetic data augmentation as proposed by Pinnis et al. (2017).

For system training, we use the *Marian* neural machine translation (NMT) toolkit by (Junczys-Dowmunt et al., 2018). We train standard NMT systems that largely follow the Transformer (Vaswani et al., 2017) *base* model configuration. The only departures from the standard configuration are the changes necessary for TLA support during training and inference. For the *Marian* toolkit, they were described in Bergmanis and Pinnis (2021a). Specifically, we employ the source-side factors using factor embeddings of dimensionality of 8 and concatenate them with subword embeddings. We also increase the delay of updates for the optimiser⁴ (from 16 to 24 batches) and set the maximum sequence length to 196 tokens. The increased sequence length accounts for longer input sequences caused by the additional TLA tokens. On the other hand, the increased optimizer delay negates the effect of the smaller effective batch size due to fewer sentences fitting in the workspace memory-based batch because of their increased maximum length. Furthermore, all models are trained using the guided alignment functionality of the *Marian* toolkit.

To train MT systems, we mostly use publicly available parallel data from the Tatoeba Challenge (Tiedemann, 2020) corpus. This constitutes 69%, 70%, and 74% of all parallel data for Latvian-Ukrainian, Estonian-Ukrainian, and Lithuanian-Ukrainian respectively. The remaining data were acquired from proprietary data sources. Data statistics are depicted in Table 1. We filtered all parallel data using parallel data filtering methods by Pinnis (2018) and then performed pre-processing, which included the following steps: normalisation of whitespaces and punctuation (e.g., quotation marks, apostrophes, hyphens, etc.), identification of non-translatable entities (e.g., e-mails, file paths, complex identifiers are replaced with placeholders), tokenisation, truecasing, synthetic

⁴ The delay of updates for the optimiser can be enabled in *Marian* using the `-optimizer-delay` parameter.

Table 1. Parallel data statistics for each language pair before and after filtering as well as after synthetic data generation.

	Before filtering	After filtering	After unknown data generation	After TLA
UK-LV	4.0M	2.3M	4.5M	8.9M
UK-ET	5.1M	4.0M	7.8M	15.5M
UK-LT	5.9M	4.7M	9.3M	18.6M

unknown data generation (Pinnis et al., 2017), byte-pair encoding (Sennrich et al., 2016b), and finally TLA.

For validation of our MT systems during training and for evaluation, we use the *dev* and *devtest* splits of FLORES-101—an evaluation benchmark specially created for low-resource language pairs (Goyal et al., 2022). We use the standard splits, which consist of 997 and 1012 sentences large validations and evaluation sets respectively that are parallel across all four languages.

Table 2. Results of automatic evaluation using SacreBLEU implementation of ChrF, BLEU and TER metrics.

		ChrF ↑ BLEU ↑ TER ↓					ChrF ↑ BLEU ↑ TER ↓		
UK-LV	eTranslation	53.7	23.0	65.4	LV-UK	eTranslation	51.4	21.2	69.8
	Google	57.0	26.9	60.7		Google	53.0	22.6	66.0
	This Work	52.2	21.2	67.3		This Work	47.6	17.5	73.0
UK-LT	eTranslation	54.0	21.4	69.0	LT-UK	eTranslation	49.1	19.1	70.7
	Google	56.6	24.4	64.5		Google	50.8	21.0	67.9
	This Work	54.5	21.7	67.7		This Work	49.4	18.8	71.0
UK-ET	eTranslation	53.6	19.4	69.2	ET-UK	eTranslation	50.6	20.8	68.6
	Google	56.2	22.0	65.7		Google	52.4	23.0	65.4
	This Work	53.5	19.4	69.2		This Work	49.5	19.5	70.5

2.1 Automatic Evaluation

We compare our systems with Google Translate⁵ and eTranslation⁶. We compare against Google Translate because, for many people, it is the go-to MT service provider when the amount of text to be translated is small. However, Google Translate is not free of charge when translation volumes exceed a certain limit. Thus we also compare against eTranslation – the MT service provider of the European Commission. eTranslation is

⁵ Accessed May 16, 2022 <https://translate.google.com>

⁶ Accessed May 16, 2022 https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-ettranslation_en

free of charge for European small and medium-sized enterprises, employees of public administrations across the European Union and public sector service providers.

Table 2 shows results of automatic evaluation using SacreBLEU (Post, 2018) implementation of three MT evaluation metrics— ChrF⁷ (Popović, 2015), BLEU⁸ (Papineni et al., 2002), and TER⁹ (Snover et al., 2006).

The automatic evaluation using ChrF, which is the most suitable metric for morphologically complex languages (Kocmi et al., 2021) such as the languages considered in this work, shows that Google Translate performs the best. Our systems compare to eTranslation in the range from marginally better for Ukrainian-Lithuanian and Lithuanian-Ukrainian directions to substantially worse for Ukrainian-Latvian. While these results do not favor our MT systems, they serve as a sanity check. Even though our systems are a one-shot attempt at developing MT systems for a set of low-resource language pairs, they are, to an extent, comparable to other publicly available alternatives.

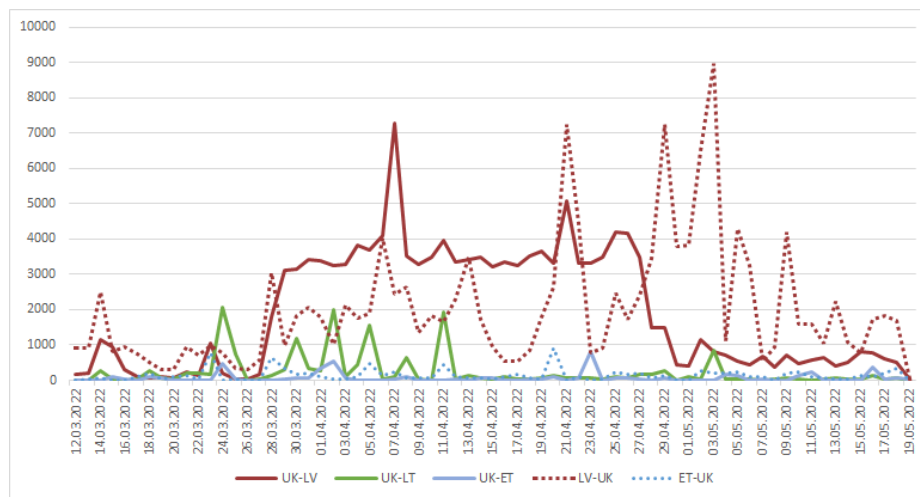


Fig. 1. The number of sentences translated daily for MT systems translating between Ukrainian and languages of Baltic states, except for the Lithuanian-Ukrainian system for which data is plotted separately (see Figure 2). Statistics are from March 11, 2022, to May 19, 2022.

3 Usage of MT Systems

We published our MT systems on March 11, 2022, which means that they have been online for more than two months at the time of writing the paper. In this section, we aim to

⁷ SacreBLEU hash: nrefs:1lcase:mixedlfff:yes lnc:6lnw:0lspace: nolversion:2.0.0

⁸ SacreBLEU hash: nrefs:1lcase:mixedlfff:no ltok:13a lsmooth:exp lversion:2.0.0

⁹ SacreBLEU hash: nrefs:1lcase:lcltok:tercomlnorm:no lpunct:yes lasian:no lversion:2.0.0

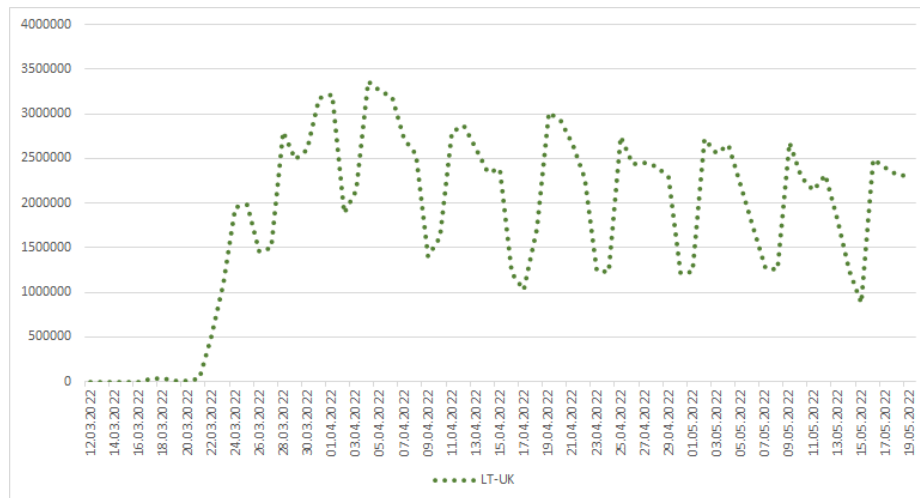


Fig. 2. The number of sentences translated daily for Lithuanian-Ukrainian MT system. Statistics are from March 11, 2022, to May 19, 2022.

analyze how our systems are used and who are their users. Figure 1 shows the number of translated sentences by each system. Due to its large translation volume, usage statistics for the Lithuanian-Ukrainian MT system are plotted separately in Figure 2. The graphs show that Estonian systems were used the least, having translated only about five thousand sentences from Ukrainian to Estonian and almost twice as much from Estonian to Ukrainian. Slightly higher usage numbers are evident for Latvian systems, which have processed more than 138 thousand and 132 thousand Latvian-Ukrainian and Ukrainian-Latvian translation sentences, respectively. Although the Ukrainian-Lithuanian system has translated only about 16 thousand sentences, the Lithuanian-Ukrainian system has had the highest demand as it has translated more than 127 million sentences.

Analyzing through what channels our systems are accessed reveals that the Latvian systems are only one-quarter of the time used by our paid clients. However, they are most often used via the Latvian language technology platform `hugo.lv`, which is popular among freelance translators and governmental organizations. As for Lithuanian systems, the users translating from Ukrainian into Lithuanian have almost exclusively used our public translation platform `translate.tilde.com`, which allows for speculation that individual users made these translation requests, most likely translating text snippets from news and social media. The system for the opposite translation direction is translating from Lithuanian into Ukrainian and has almost entirely been used via `Tilde Web Translation Widget`. To understand the 127 million sentences large volume of translated sentences, we inspect the distribution of translated sentences by their source website (see Figure 3). All websites using this MT system are related to the Lithuanian government. The top websites are `uzt.lt` and `ldb.lt`, which are services of the employment agency of Lithuania, `paslaugos.vilnius.lt`, which is the Vilnius City Council services' page and `socmin.lrv.lt`, which is the Ministry of Social

Security and Labor of the Republic of Lithuania. This analysis reveals that, at least as far as the usage of the Lithuanian-Ukrainian MT system is concerned, even if just a little, our work has helped the people in need to access help and social services.

It is also important to note that the difference of the usage levels for Latvian and Lithuanian systems can be explained with how the systems are used in Latvia and Lithuania. In Lithuania, the Lithuanian-Ukrainian system is (mostly) used to translate governmental websites. Whenever a user (a citizen, a refugee, or a tourist) accesses a certain page in a website, its content is translated by the MT system. This generates high numbers of translation requests. However, this method allows to provide instant multilinguality in a website regardless of which page a user wants to see. In Latvia, the systems are mostly used by translators and public service officials in post-editing scenarios. This means that different from Lithuania where we can grasp a rough estimate of how many end-users consume the translations, in Latvia we only know how many unique sentences were translated to create content in a different language. We cannot estimate how many end-users might have consumed that content. However, the volume is still substantial for post-editing scenarios.

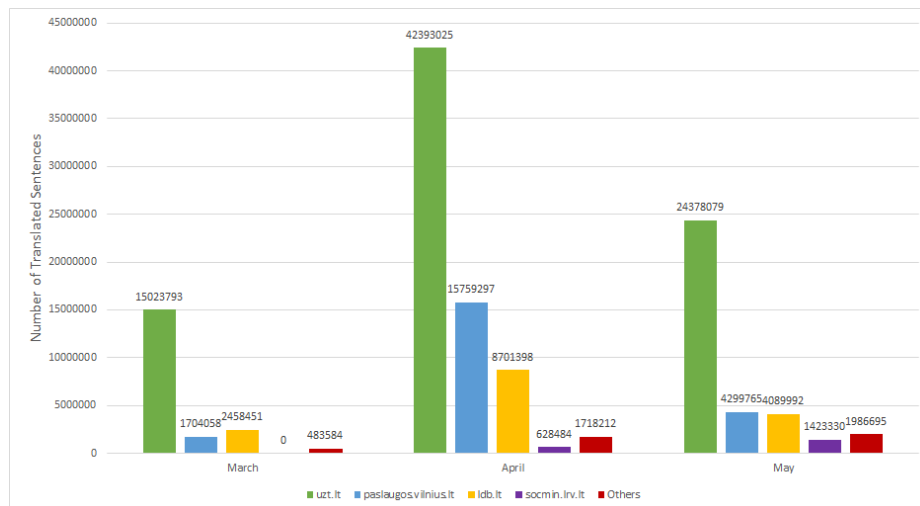


Fig. 3. The number of sentences translated by the top users of the Lithuanian-Ukrainian MT system. Statistics are from March 11, 2022, to May 19, 2022.

4 Step Forward

4.1 Lithuanian-Ukrainian MT System with Back-translated Data

In Section 3, we established that the Lithuanian-Ukrainian MT system is used the most as it has translated 127 million sentences helping Ukrainians in Lithuania to find jobs

Table 3. Results of automatic evaluation using the SacreBLEU implementation of ChrF, BLEU and TER metrics for Lithuanian-Ukrainian MT systems. *This Work’s BT System* refers to the system developed using back-translated data, while *This Work’s Baseline* refers to MT systems described in Section 4.1.2. * denotes that the result difference between this and the system trained on back-translated data is statistically significant according to SacreBLEU’s paired bootstrap re-sampling test.

	ChrF ↑	BLEU ↑	TER ↓
eTranslation	49.1*	19.1	70.7
Google Translate	50.8*	21.0	67.9
This Work’s Baseline	49.4*	18.8	71.0
This Work’s BT System	50.0	19.3	70.7

and access social services. Besides, unlike the Latvian-Ukrainian system, which is primarily used in post-editing scenarios, the translations of the Lithuanian-Ukrainian system reach its users without the supervision of professional translators. Therefore, we aim to deliver better technology where the people use it the most and retrain our Lithuanian-Ukrainian MT system.

4.1.1 Machine Translation Systems We use the Ukrainian-Lithuanian MT system to create synthetic parallel data by back-translation (Sennrich et al., 2016a) of monolingual Ukrainian data. For data sources, we use the 2008 to 2021 News Crawl¹⁰ corpus provided by the Machine Translation Group at the University of Edinburgh and the RSS News, Newscrawl, and Wikipedia corpora¹¹ collected by the University of Leipzig (Goldhahn et al., 2012). We also use the Ukrainian side of the Ukrainian-English Wikimedia (Tiedemann, 2012), TED 2020 (Reimers and Gurevych, 2020), and OpenSubtitles v2018 (Lison and Tiedemann, 2016) corpora from Opus¹² (Tiedemann, 2012). Altogether these corpora amount to around 11.4 million sentences.

As before, we continue by using the synthetic data augmentation (Pinnis et al., 2017), which nearly doubles the number of sentences to about 21 million. We then translate this data into Lithuanian and use parallel data filters by Pinnis (2018) to get rid of noisy and poor quality translations, which leaves us with around 19 million back-translated sentences. Finally, we add this data to the data we used to train the baseline system (see Section 4.1.2) to obtain a total of about 37.8 million sentences. We then use the same configuration as described in Section 4.1.2 with an exception that we increase optimizer delay from 24 to 64.

4.1.2 Automatic Evaluation Table 3 shows a comparison of the baseline MT systems from Section and the newly created Lithuanian-Ukrainian MT system. The new system achieves the second best results, conceding only to Google Translate, which is still 0.8

¹⁰ <https://data.statmt.org/news-crawl/uk>

¹¹ <https://wortschatz.uni-leipzig.de/en/download/Ukrainian>

¹² <https://opus.nlpl.eu/>

ChrF points better. However, the results also show that using back-translated data helps to yield statistically significant improvements in translation quality over the other two baselines.

Table 4. Examples of Ukrainian toponyms and their translations previously represented via their pronunciation in Russian.

UK	EN		LV		LT	
	New	Obsolete	New	Obsolete	New	Obsolete
Київ	Kyiv	Kiev	Kijiva	Kijeva	Kyjivas	Kyjevas
Харків	Kharkiv	Kharkov	Harkiva	Harkova	Charkivas	Charkovas
Одеса	Odessa	Odesa	Odesa	-	Odesa	-
Дніпро	Dnipro	Dnipropetrovsk	Dnipro	Dņepro	Dnipras	Dniepras
Донецьк	Donetsk	-	Donecka	Doņecka	Doneckas	-
Запоріжжя	Zaporizhzhia	-	Zaporižja	Zaporožje	Zaporižė	Zaparožė
Львів	Lviv	Lvov	Lviva	Lvova	Lvivas	Lvovas
Кривий Ріг	Kryvyi Rih	-	Krivijriha	Krivojroga	Kryvyj Rihas	-
Миколаїв	Mykolaiv	Nikolaev	Mikolajiva	Mikolajeva	Mykolajevas	-
Луганськ	Luhansk	Lugansk	Luhanska	Luganska	Luhanskas	-
Макіївка	Makiivka	Makiyivka	Makijivka	-	Makijivka	-
Вінниця	Vinnytsia	-	Vinnica	Vinņica	Vinica	-
Чернігів	Chernihiv	Chernigov	Černihiva	Čerņigova	Čerņyhivas	Čerņyhovas
Чернівці	Chernivtsi	Makiyivka	Černivci	Černovci	Černivciai	-
Горлівка	Horlivka	-	Horlivka	Gorlovka	Horlivka	-
Кам'янське	Kamianske	-	Kamjanska	Kamjanske	Kamianskė	-
Тернопіль	Ternopil	Tarnopol	Ternopiļa	Ternopole	Ternopilis	Ternopolis

4.2 Ukrainian Toponym Translation

Historically Ukrainian toponyms in the languages of the Baltic states have been introduced via Russian. Thus traditionally, Latvian and Lithuanian representations of Ukrainian toponyms have leaned on the conventions of Russian pronunciation. Traditions, however, are subject to cultural changes, as exemplified by the decommunization of Ukrainian toponymy after the collapse of the USSR and the proclamation of independent Ukraine (Demska and Levchuk, 2020). Likewise, shifts in geopolitical allegiances can also be a decisive factor in changing language customs. Here, the example is the departure from the Russian-based representations of Ukrainian toponyms in Latvian to favour Ukrainian-based pronunciation. Since 2014 when the Russian Federation started military aggression against Ukraine, the expert committee of the Latvian State Language Centre has twice pushed for changes in the Latvian language representations of Ukrainian city names – first in 2017¹³ and then in 2019.¹⁴ The final decision to offi-

¹³ Accessed May 5, 2022 <https://www.vestnesis.lv/op/2017/208.18>

¹⁴ Accessed May 5, 2022 <https://www.vestnesis.lv/op/2019/42.37>

Table 5. Examples of Ukrainian-Latvian and Ukrainian-Lithuanian toponym translation with and without terminology integration (incorrect translations of toponyms are underlined).

Ukrainian-Latvian	
Source:	Київ, Харків, Одеса, Дніпро і Донецьк мають п'ять найбільших українських міст.
Without terms:	<u>Kije</u> va, <u>Harkov</u> a, Odesa, <u>Dņep</u> ra un <u>Doņeck</u> a ir piecas lielākās Ukrainas pilsētas.
With terms:	Kijīva, Harkīva, Odesa, Dņipro un Donecka ir piecas lielākās Ukrainas pilsētas.
Ukrainian-Lithuanian	
Source:	Ми доставляємо товари в Запоріжжя, Львів, Чернігів та Тернопіль.
Without terms:	Mes pristatome prekes į <u>Zaporož</u> ę, <u>Lvov</u> ą, <u>Černig</u> ą ir Ternopilį.
With terms:	Prekes pristatome į Zaporizę, <u>Lvov</u> ą, Černyhivą ir Ternopilį.

cially stop using Russian-based representations of Ukrainian city names was reached on March 9, 2022,¹⁵ only two weeks after the Russian invasion of Ukraine.

Although these swift decisions reflect the political climate and the sentiment of the people, these changes hardly have had time to reach the training data of data-driven natural language processing tools. So we take advantage of our MT system's dynamic terminology integration capability and approach the problem of Ukrainian toponym translation as a terminology integration task. Specifically, we prepare toponym glossaries (see Table 4) mapping both the new and obsolete terms to their new and preferred translations. Before translating, we compare the stemmed version of each word in the sentence against the stemmed Ukrainian toponyms in the glossary and annotate them with their preferred translation if we find one. Then, we pass the annotated sentence to a system that is trained with TLA and can use the annotations to translate and inflect the toponym according to the sentence context. For more details, refer to Bergmanis and Pinnis (2021b) and Bergmanis and Pinnis (2021a).

Terminology integration using TLA applies soft constraints on an NMT model. Contrary to methods that apply hard constraints, e.g., constrained decoding (Post and Vilar, 2018), this enables the NMT model to have flexibility in how the annotations are used. The NMT system can freely decide on the most suitable inflected form for the given morphosyntactic context. However, this also means that in some cases, the NMT model can choose to ignore the annotations if there is a stronger internal signal for a different lexical choice. Table 5 shows two examples where Ukrainian toponyms are translated from Ukrainian into Latvian and Lithuanian. The example shows that terminology integration improves toponym translation quality for most cases except for one example, 'Львів', was translated using the obsolete variant. Nevertheless, we believe that soft constraints are more appropriate for morphologically rich languages. There is

¹⁵ Accessed May 5, 2022 <https://www.vestnesis.lv/op/2022/67.4>

room for future work to reduce cases where the NMT model decides not to rely on the annotations.

5 Conclusions and Discussion

We examined the quality, usage patterns and translation volume of six low-resource MT systems translating between the Ukrainian language and each of the official languages of the Baltic states. Although the translation quality analysis revealed that our systems are no better than the other publicly available alternatives, the MT usage statistics showed that the general public nevertheless uses some of our MT systems. Meanwhile, other MT systems are integrated into Lithuania's governmental websites or used by government translators in Latvia.

We found that the different approaches to MT integration in public services have led to vastly different volumes of translation requests. In Lithuania, whenever a user accesses a certain page on a website, the MT system translates its content, generating many translation requests. This method provides flexible and instant multilingualism on a website regardless of which page a user wants to access. In Latvia, the systems are used mainly by translators and public service officials in MT post-editing scenarios. While this approach generates fewer translation requests, it also limits what content users can access in their native language.

Knowing which systems are used most actively, we revisit them to improve their quality. Because of the different ways systems are integrated, in contrast to the Latvian-Ukrainian system, the translations of the Lithuanian-Ukrainian system reach users without being checked by professional translators. Motivated by this finding, we retrained the Lithuanian-Ukrainian MT system using nearly twenty million sentences of back-translated data, which allowed us to, as measured by automatic metrics, outperform eTranslation and close the gap with Google Translate. Finally, we cast the Ukrainian toponym translation as a terminology integration task and show how to dynamically solve the changing and divergent spelling of place names when systems are deployed.

There are options for future work to improve the MT between the official languages of Baltic states and Ukrainian beyond the quality achieved within this work. One, evident from Table 1, is to obtain more high-quality data. Indeed, the amount of training data after filtering ranges from two to nearly five million parallel sentences, which is not much compared to other European language pairs. Another potential avenue for future work is to train multilingual MT models (Dabre et al., 2020) translating from many source languages to one target language. In such a setup, including one resource-rich language pair, such as English-Ukrainian, could help via means of transfer-learning (Kocmi, 2019), or at the very least, as a form of regularization (Neubig and Hu, 2018).

While this work provides a novel analysis of the MT usage in Baltic states to address language barriers rising from a refugee crisis, there are other similar efforts to use language technology to aid people displaced by the Russian war in Ukraine. One such effort is ÚFAL for Ukraine by Charles University in Prague, which offers an MT system for Czech-Ukrainian, Charles Translator for Ukraine. Their MT system builds on the previous work on Czech-English MT (Popel et al., 2020) and can be accessed

on the web,¹⁶ via an android app,¹⁷ as well as in the form of chatbots for Telegram and Messenger and other messaging services.¹⁸

Acknowledgements

The research has been supported by the European Regional Development Fund within the research project “AI Assistant for Multilingual Meeting Management” No. 1.1.1.1/19/A/082.

References

- Bergmanis, T., Pinnis, M. (2021a). Dynamic terminology integration for COVID-19 and other emerging domains, *Proceedings of the Sixth Conference on Machine Translation*, Association for Computational Linguistics, Online, pp. 821–827.
- Bergmanis, T., Pinnis, M. (2021b). Facilitating terminology translation with target lemma annotations, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Association for Computational Linguistics, Online, pp. 3105–3111.
- Cosgrove, J. (2020). The Russian invasion of the Crimean peninsula, 2014-2015: A post-cold war nuclear crisis case study, *Technical report*, Applied Physics Laboratory, Johns Hopkins University.
- Dabre, R., Chu, C., Kunchukuttan, A. (2020). A survey of multilingual neural machine translation, *ACM Computing Surveys (CSUR)* **53**(5), 1–38.
- Demska, O., Levchuk, P. (2020). The urbanonimia of Ukraine in the context of decommunization, *Cognitive Studies/ Études cognitives* (20).
- Goldhahn, D., Eckart, T., Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig corpora collection: From 100 to 200 languages, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, pp. 759–765.
- Goyal, N., Gao, C., Chaudhary, V., Chen, P.-J., Wenzek, G., Ju, D., Krishnan, S., Ranzato, M., Guzmán, F., Fan, A. (2022). The FLORES-101 evaluation benchmark for low-resource and multilingual machine translation, *Transactions of the Association for Computational Linguistics* **10**, 522–538.
- Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Fikri Aji, A., Bogoychev, N., Martins, A. F. T., Birch, A. (2018). Marian: Fast Neural Machine Translation in C++, *Proceedings of ACL 2018, System Demonstrations*, Association for Computational Linguistics, Melbourne, Australia, pp. 116–121.
- Kocmi, T. (2019). Exploring benefits of transfer learning in neural machine translation.
- Kocmi, T., Federmann, C., Grundkiewicz, R., Junczys-Dowmunt, M., Matsushita, H., Menezes, A. (2021). To ship or not to ship: An extensive evaluation of automatic metrics for machine translation, *Proceedings of the Sixth Conference on Machine Translation*, Association for Computational Linguistics, Online, pp. 478–494.

¹⁶ Accessed July 14, 2022 <https://lindat.cz/translation>

¹⁷ Accessed July 14, 2022 <https://play.google.com/store/apps/details?id=cz.cuni.mff.ufal.translator>

¹⁸ Accessed July 14, 2022 <https://github.com/martin-majlis/uk-cs-translation-bot>

- Lison, P., Tiedemann, J. (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, European Language Resources Association (ELRA), Portorož, Slovenia, pp. 923–929.
- Neubig, G., Hu, J. (2018). Rapid adaptation of neural machine translation to new languages, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Brussels, Belgium, pp. 875–880.
- Papineni, K., Roukos, S., Ward, T., Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation, *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318.
- Pinnis, M. (2018). Tilde's parallel corpus filtering methods for WMT 2018, *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, Association for Computational Linguistics, Belgium, Brussels, pp. 939–945.
- Pinnis, M., Krišlauks, R., Deksne, D., Miks, T. (2017). Neural machine translation for morphologically rich languages with improved sub-word units and synthetic data, *International conference on text, speech, and dialogue*, Springer, pp. 237–245.
- Popel, M., Tomkova, M., Tomek, J., Kaiser, Ł., Uszkoreit, J., Bojar, O., Žabokrtský, Z. (2020). Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals, *Nature communications* **11**(1), 1–15.
- Popović, M. (2015). chrF: character n-gram f-score for automatic mt evaluation, *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pp. 392–395.
- Post, M. (2018). A call for clarity in reporting bleu scores, *Proceedings of the Third Conference on Machine Translation: Research Papers*, pp. 186–191.
- Post, M., Vilar, D. (2018). Fast lexically constrained decoding with dynamic beam allocation for neural machine translation, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1314–1324.
- Reimers, N., Gurevych, I. (2020). Making monolingual sentence embeddings multilingual using knowledge distillation, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics.
- Sennrich, R., Haddow, B., Birch, A. (2016a). Improving neural machine translation models with monolingual data, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Berlin, Germany, pp. 86–96.
- Sennrich, R., Haddow, B., Birch, A. (2016b). Neural machine translation of rare words with subword units, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, Association for Computational Linguistics, Berlin, Germany.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J. (2006). A study of translation edit rate with targeted human annotation, *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pp. 223–231.
- Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, pp. 2214–2218.
- Tiedemann, J. (2020). The Tatoeba translation challenge—realistic data sets for low resource and multilingual MT, *Proceedings of the Fifth Conference on Machine Translation*, pp. 1174–1182.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017). Attention Is All You Need, *Advances in Neural Information Processing Systems*, pp. 5998–6008.

Received July 18 , accepted July 28, 2022