

BLEU Skies for Endangered Language Revitalization: Lemko Rusyn and Ukrainian Neural AI Translation Accuracy Soars

Petro Orynych^[0000-0003-3094-9156]

Orynych.com, Edgewater MD 21037, USA
p@orynych.com

Abstract. Accelerating global language loss, associated with elevated incidence of illicit substance use, type 2 diabetes, binge drinking, and assault, as well as sixfold higher youth suicide rates, poses a mounting challenge for minority, Indigenous, refugee, colonized, and immigrant communities. In environments where intergenerational transmission is often disrupted, artificial intelligence neural machine translation systems have the potential to revitalize heritage languages and empower new speakers by allowing them to understand and be understood via instantaneous translation. Yet, artificial intelligence solutions pose problems, such as prohibitive cost and output quality issues. A solution is to couple neural engines to classical, rule-based ones, which empower engineers to purge loanwords and neutralize interference from dominant languages. This work describes an overhaul of the engine deployed at LemkoTran.com to enable translation into and out of Lemko, a severely endangered, minority lect of Ukrainian genetic classificability indigenous to borderlands between Poland and Slovakia (where it is also referred to as Rusyn). Dictionary-based translation modules were fitted with morphologically and syntactically informed noun, verb, and adjective generators fueled by 877 lemmata together with 708 glossary entries, and the entire system was riveted by 9,518 automatic, codification-referencing, must-pass quality-control tests. The fruits of this labor are a 23% improvement since last publication in translation quality into English and 35% increase in quality translating from English into Lemko, providing translations that outperform every Google Translate service by every metric, and score 396% higher than Google's Ukrainian service when translating into Lemko.

Keywords: Lemko, Neural Machine Translation, Language Revitalization.

Please cite as: Orynych, P. (2023). BLEU Skies for Endangered Language Revitalization: Lemko Rusyn and Ukrainian Neural AI Translation Accuracy Soars. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCI 2023. Lecture Notes in Computer Science(), vol 14051. Springer, Cham. https://doi.org/10.1007/978-3-031-35894-4_10

This version of the contribution has been accepted for publication after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at https://doi.org/10.1007/978-3-031-35894-4_10. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>.

1 Introduction

1.1 The Problem

Languages are being lost at a rate of at least one per calendar quarter, with such loss set to triple by 2062, and increase fivefold by 2100, affecting over 1,500 speaker communities [1, pp. 163 and 169]. Such outcomes are associated with elevated incidence of illicit substance use [2, p. 179], type 2 diabetes [3], binge drinking, and assault [4], as well as sixfold higher youth suicide rates when fewer than half of community members have language knowledge [5].

A recent study in the United States found that Indigenous language use has positive effects on health, regardless of proficiency level [6]. An experiment on speakers in Poland has found that use of Lemko moderates emotional, behavioral, and depressive symptoms stemming from cognitive availability of trauma [7].

Artificial intelligence machine translation might be of service in spreading the aforementioned protective effects to heritage speakers by revitalizing dying and Sleeping languages [8, p. 577]. For example, new speakers might produce correct text instantaneously and enjoy reading comprehension using automatic machine translation devices as an aid until full, independent fluency is achieved.

1.2 System Under Study

Language. Lemko is a definitively to severely endangered [9, pp. 177–178] East Slavic lect of southwestern Ukrainian genetic classificability [10, p. 52; 11, p. 39] indigenous to borderlands between the Republic of Poland and Slovak Republic; some have referred to it as Rusyn [11, p. 39; 12].

Eastern boundaries. A unique isogloss differentiating Lemko to the East is fixed paroxytonic (penultimate syllable) stress, a feature shared with Polish and Eastern Slovak dialects [10, pp. 161–162 and 972–973; 11, p. 50; 13, pp. 70–73], making its extent in Eastern Slovakia at least to the Laborec River, with a transitional zone extending thereafter [13, p. 70; 11, p. 50]. Meanwhile in Poland, the historical extent of Lemko reaches at least the Oślawica or Wisłok rivers, with a transitional zone beyond them [11, p. 50].

Western boundaries. The historical western boundaries of Lemko are the Poprad and Dunajec rivers [14, p. 459].

Locale. Ancestral villages of native speakers whose interviews comprise the corpus are found within the current administrative borders of today's Lesser Poland Province, whose capital is Cracow.

Table 1. Ancestral villages of native speakers interviewed in corpus material.

Lemko name	Transliteration	Polish name	County Seat	Commune Seat
Ізбы	Izbŷ	Izby	Gorlice	Uście Gorlickie
Гладъшів	Gładŷšiv	Gładyszów	Gorlice	Uście Gorlickie
Чорне	Čorne	Czarne	Gorlice	Sękowa
Долге	Dolhe	Długie	Gorlice	Sękowa
Білцарьова	Bilcar'ova	Binczarowa	Nowy Sącz	Grybów
Фльоринка	Fl'orynka	Florynka	Nowy Sącz	Grybów
Чырна	Čŷrna	Czyrna	Nowy Sącz	Krynica-Zdrój

2 State of the Art

Last year, the world's first quality evaluation results were published for machine translations into Lemko: BLEU 6.28, which was nearly triple that of Google Translate's Ukrainian service¹ (BLEU 2.17) [15, p. 570]. The year before, my colleagues and I had published and presented the world's first results for Lemko to English machine translation: BLEU 14.57 [16].

The engine has been deployed and made freely available at the universal resource locator <https://www.LemkoTran.com>, where a transliteration engine has been in service since the autumn of 2017. The translation engine was first alluded to in print by Drs. Scherrer and Rabus in the Cambridge University Press journal *Natural Language Engineering* in 2019 [17].

3 Materials and Methods

3.1 Materials

The experiment was performed on a bilingual corpus comprising Lemko Cyrillic transcripts and English translations of interviews with survivors and children of forced re-settlements from ancestral lands in Poland. The transcripts and their translations² were aligned across 3,267 segments, with Microsoft Word providing a Lemko source word count of 68,944 and an English target word count of 81,188.

¹ Disclosure: I work as a paid Ukrainian, Polish, and Russian translation quality control specialist for the Google Translate project. My client's headquarters are in San Francisco, California.

² I was hired to produce the transcripts and translate them by the John and Helen Timo Foundation of Wilmington, Delaware, who then donated the work products to my scientific research and development endeavors.

Sources of truth included the dictionaries of Jarosław Horoszczak [18], Petro Pyrtej [19], Ihor Duda [20], and Janusz Rieger [21], as well as the grammars of Henryk Fontański and Mirosława Chomiak [22] and Petro Pyrtej [23].

3.2 Methods

Engine Upgrades. For this experiment, the engine deployed at `LemkoTran.com` was fitted with newly built generators informed by part of speech, grammatical case, and number for the purpose of producing grammatically and syntactically appropriate translations for 1,585 dictionary entries, about half of which do not inflect in Polish or Lemko, allowing for simple substitution.

Quality Assurance Tests. Quality was ensured by 9,518 tests cross-referenced when feasible with the Lemko codifications, grammars, and dictionaries listed above under *Materials*. The tests themselves assert that the system translates given utterances in the desired manner.

Table 2. System vocabulary.

Description	Quantity
Noun stem	414
Verb stem	296
Adjective stem	167
Pronoun, personal	87
Pronoun, other	178
Numeral	86
Other dictionary entries	357
Total	1,585

Rule-Based Machine Translation (RMBT). Text was given a Lemko or Polish look and feel by replacing character sequences, and especially inflectional endings.

Table 3. Example character sequence replacements.

Polish Sequence	Lemko Sequence	Position
ować	uwaty	Final
i ami	i amy	Final
a ja	a jut	Final
ze	zo	Initial
pod	pid	Initial

Translation Quality Scoring. Translation quality was measured per industry standard metrics using the default settings of the SACREBLEU tool invented at Amazon Research

by Matt Post [24]. For the sake of comparability, Polish was rendered in Lemko Cyrillic in the same way as the last experiment [15, p. 573].

Bilingual Evaluation Understudy (BLEU). This n -gram-based metric has enjoyed wide currency for decades. It was developed in the United States at the IBM T. J. Watson Research Center with support from the Defense Advanced Research Projects Agency (DARPA) and monitoring by the United States Space and Naval Warfare Systems Command (SPAWAR) [25].

Translation Edit Rate (TER). This metric reflects the number of edits necessary for output to semantically approach a correct translation, aiming to be more tolerant of phrasal shifts than BLEU and other n -gram-based metrics. It is determined by dividing a calculation of edit distance between a hypothesis and a reference by average reference wordcount. Its development in the United States was also supported by DARPA [26].

Character n -gram F -score (CHRF). This European metric been shown to correlate very well with human judgments and even outperform both BLEU and TER [27].

4 Results and Discussion

The experimental system, LemkoTran.com, outperformed every Google Translate service by every metric. English to Lemko translation BLEU quality scores improved 35% in comparison with last published results [15], producing results four times better than Google Translate’s next-best offering, its Ukrainian service. Meanwhile, Lemko to English translation quality improved by 23% since last published results [16], achieving BLEU scores 16% higher than the best obtained by Google Translate, which automatically recognized Lemko as Ukrainian 76% of the time, as Russian 16% of the time, and as Belarusian 6% of the time.

4.1 English to Lemko Translation Quality

Scores. The engine deployed at LemkoTran.com bested Google Translate by every metric when translating from English into Lemko. The next-highest scoring system in the experiment was either the output of Google Translate’s Ukrainian service (using the BLEU or CHRF metrics) or that of its Polish service (using the TER metric).

BLEU. The translation quality of the system deployed at LemkoTran.com as measured by the most widespread BLEU metric rose to 8.48, a 35% improvement on results last published in 2022 [15], and now quadruple Google Translate’s highest score.

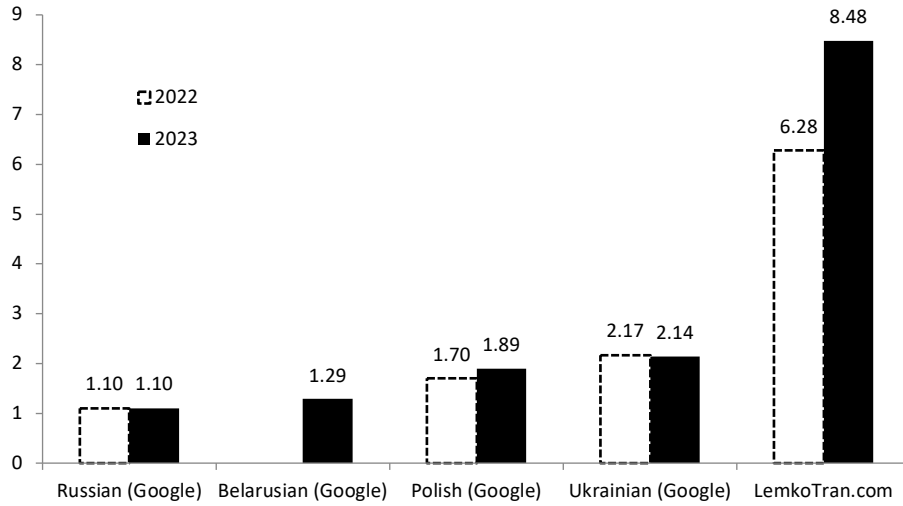


Fig. 1. English to Lemko translation quality as measured by Bilingual Evaluation Understudy (BLEU) score, Google Cloud Neural Machine Translation (NMT) services versus LemkoTran.com. The higher, the better.

CHRF. The LemkoTran.com engine achieved the best English to Lemko character n -gram f -score (CHRF 37.30), which is 37% higher than the next best, Google Translate's Ukrainian service. Meanwhile, Google Translate's Russian service scored higher than its Polish and Belarusian counterparts when measured against the Lemko corpus by this metric.

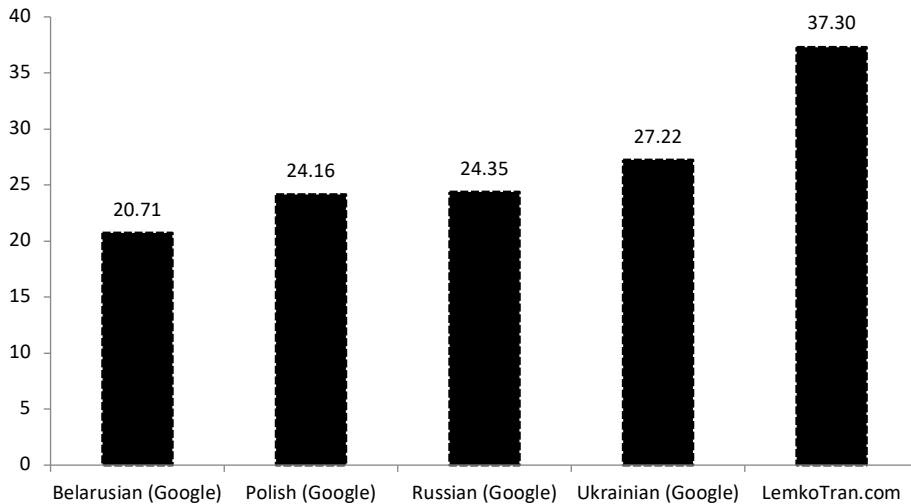


Fig. 2. English to Lemko translation quality as measured by character n -gram F-score (CHRF)

score, Google Cloud Neural Machine Translation (NMT) versus the experimental system LemkoTran.com. The higher, the better.

TER. The LemkoTran.com engine achieved the best English to Lemko Translation Edit Rate (TER), scoring 81.33. Google Translate's Polish service scored second best, followed closely by its Ukrainian one.

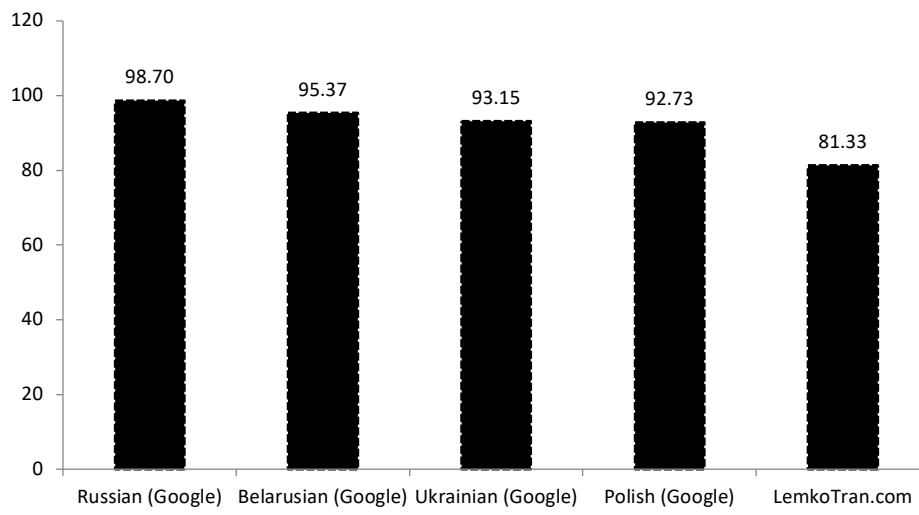


Fig. 3. English to Lemko Translation Edit Rate (TER), Google Cloud Neural Machine Translation (NMT) versus LemkoTran.com. The lower, the better.

Samples. Output from the translation systems when fed English is given below.

Table 4. Comparisons of translation hypotheses for English input.

Input	Our children were smart too. But where were they supposed to study?		
Description	Output	Transliteration	Quality Scores
Lemko reference (native speaker)	В нас діти тіж были мудры, але де мали ся вчыти?	V nas dity tiž bŷly mudrŷ, ale de maly sja včŷty?	BLEU 100 CHRF2 100 TER 0
Translation into Lemko by LemkoTran.com	Нашы діти тіж были мудры. але де мали ся вчыти?	Našŷ dity tiž bŷly mudrŷ. ale de maly sja včŷty?	BLEU 58.34 CHRF2 79.03 TER 27.27
Google Translate (control)	Translation into Ukrainian	Наші діти теж були розумними. Але де вони мали вчитися?	BLEU 4.41 CHRF2 25.80 TER 72.73
	Translation into Russian	Наші діти теж були умными. Но где им было учиться?	BLEU 3.71 CHRF2 16.95 TER 90.91
	Translation into Polish	Наше дзеці теж били мондре. Але гдзе мелі се учиць?	BLEU 3.12 CHRF2 13.84 TER 100
	Translation in Belarusian	Разумныя былі і нашы дзеці. Але дзе яны павінны былі вучыцца?	BLEU 3.09 CHRF2 12.83 TER 100

Table 5. Comparisons of translation hypotheses for English input.

Input	And generally speaking, Lemkos in Poland don't have a leader, so to speak, who would say something.		
Description	Product	Transliteration	Quality Scores
Lemko reference (native speaker)	А вогулі Лемкы в Польшы не маюť такого, же так повім, такого лідера, котрий бы штоси повіл.	A voguli Lemkŷ v Pol'sščŷ ne majut takoho, že tak povim, takoho lidera, kotryj bŷ štosy povil.	BLEU 100 CHRF2 100 TER 0
Translation into Lemko by LemkoTran.com	І генеральні Лемкы в Польшы не маюť лідера, же так повім, котрий бы штоси повіл.	I heneral'ni Lemkŷ v Pol'sščŷ ne majut lydera, že tak povim, kotryj bŷ štosy povil.	BLEU 55.58 CHRF2 65.32 TER 29.41
Google Translate (control)	Translation into Polish	І генеральне Лемкове в Польсце не майон лідера, же так повем, ктури би цось поведзял.	BLEU 9.26 CHRF2 29.29 TER 82.35
	Translation into Ukrainian	І взагалі, лемки в Польщі не маюť лідера, так би мовити, який би щось сказав.	BLEU 5.15 CHRF2 26.56 TER 82.35
	Translation into Russian	И вообще, у лемков в Польше нет, так сказать, лидера, который бы что-то сказал.	BLEU 2.96 CHRF2 25.87 TER 88.24
	Translation into Belarusian	І ўвогуле лэмкі ў Польшы ня маюць лідэра, так бы мовіць, які б нешта сказаў.	BLEU 2.72 CHRF2 18.05 TER 94.12

4.2 Lemko to English Translation

Scores. For every metric, the engine deployed at LemkoTran.com outperformed Google Translate, for which translation as if from Standard Ukrainian was always second best, followed by it automatically detecting the source language, then translating as if from Belarusian, and then Polish, with Russian always coming in last place. Google Translate recognized Lemko as Ukrainian 76% of the time, as Russian 16% of the time, as Belarusian 6% of the time, and as sundry languages using Cyrillic alphabets (e.g. Mongolian) the rest of the time.

BLEU. LemkoTran.com scored BLEU 17.95 when translating into English, a 23% improvement on last published results of BLEU 14.57, and 16% higher than Google Translate’s Ukrainian service’s score of BLEU 15.43.

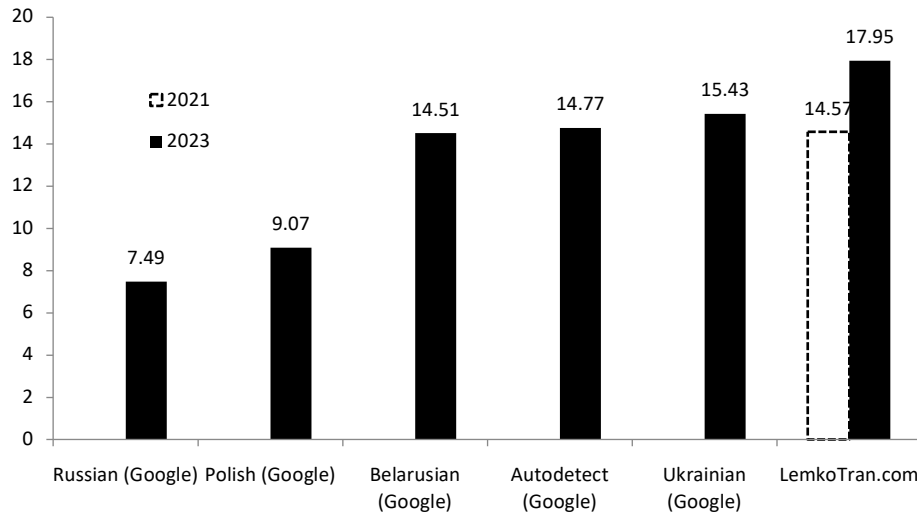


Fig. 4. Lemko to English translation quality as measured by Bilingual Evaluation Understudy (BLEU) score, Google Cloud Neural Machine Translation (NMT) services versus the experimental system LemkoTran.com. The higher, the better.

CHRF. The engine deployed at LemkoTran.com achieved a character n -gram f-score (CHRF) of 45.89 when translating into English, which was 5% better than the score of Google Translate’s Ukrainian service.

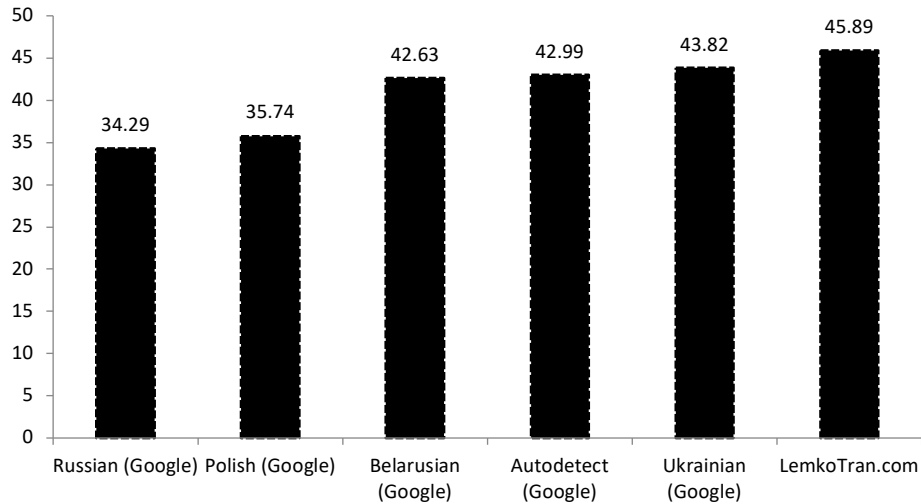


Fig. 5. Lemko to English translation quality as measured by character n -gram F-score (CHRF) score, Google Cloud Neural Machine Translation (GNMT) versus the experimental system LemkoTran.com. The higher, the better.

TER. LemkoTran.com scored a Translation Edit Rate (TER) of 70.38 translating into English, which was 7% better than the score of Google Translate’s Ukrainian service.

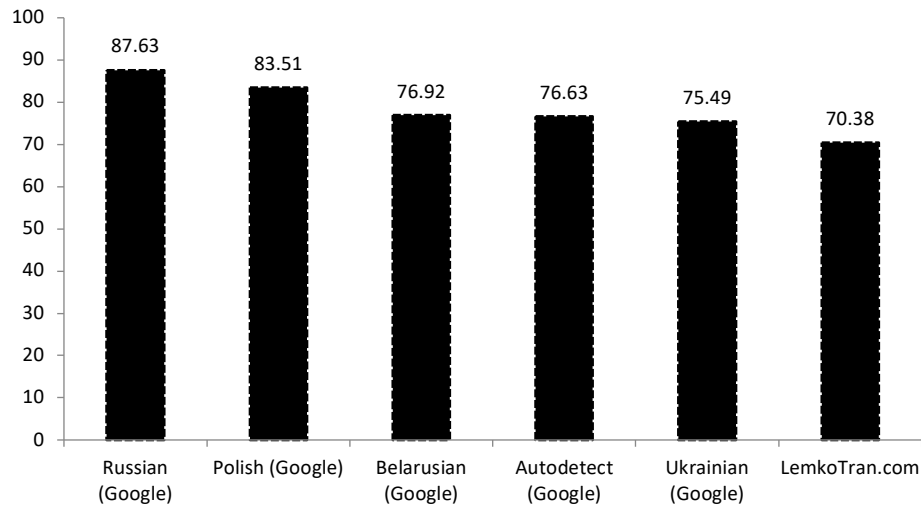


Fig. 6. Lemko to English Translation Edit Rate (TER), Google Cloud Neural Machine Translation (GNMT) versus the experimental system LemkoTran.com. The lower, the better.

Samples. Output from the translation systems when fed English is given below.

Table 6. Comparisons of translation hypotheses for Lemko input.

Description	Product	Quality Scores
Input transcription of Lemko spoken by a native speaker	Як розділяме яззыкы, то мала-м контакт з польськым, то не было так, же пішла-м до школы без польського, бо зме мали сусідів Поляків.	n/a
Transliteration	Jak rozdiljame jazÿkÿ, to mala-m kontakt z pol'skÿm, to ne bÿlo tak, že pišla-m do školÿ bez pol'skoho, bo zme maly susidiv Poljakiv.	n/a
Reference translation by a bilingual professional	When it comes to separating languages, I had contact with Polish. It wasn't like I started school without knowing Polish because we had Polish neighbors.	BLEU 100 CHR2 100 TER 0
Translation from Lemko by the system at LemkoTran.com	When we separate languages, I had contact with Polish, it wasn't like I went to school without Polish, because we had Polish neighbors.	BLEU 45.84 CHR2 69.60 TER 32.00
from Ukrainian (autodetected with 92% confidence)	As we divide the languages, then I had contact with Polish, then it was not like that, and I went to school without Polish, because I had Poles as neighbors.	BLEU 15.87 CHR2 54.38 TER 72.00
from Belarusian	As we separate the languages, then I had little contact with Polish, then it was not like that, but I went to school without Polish, because we had few Polish neighbors.	BLEU 11.76 CHR2 58.92 TER 68.00
from Russian	As we spread languages, then there was little contact with Polish, then it wasn't like that, but I went to school without Polish, for the snakes were sucid in Polyakiv.	BLEU 6.87 CHR2 42.66 TER 92.00
from Polish	As I spread the language, I have little contact with the Polish language, it wasn't like that I went to school without Polish, because I will change my little Polish language.	BLEU 5.02 CHR2 45.35 TER 84.00

5 Conclusion

Coupling morphologically and syntactically informed generators to neural engines can improve machine translation quality by at least a third, while also having the side benefit of empowering engineers to purge loanwords and counteract other dominant-language interference, as well as ensure compliance with standards, such as codifications of minority languages. Quality-score glass ceilings imposed by the imperfections inherent to artificial intelligence models can also be shattered through sound engineering. For Lemko, as well as fellow low-resource, Indigenous minority languages, the sky is now the limit for translation quality, as well as revitalization revolutions just over the horizon.

Acknowledgements. I would like to thank Dr. Ming Qian of Charles River Analytics for the inspiration to conduct this experiment, Michael Decerbo of Raytheon BBN Technologies and Dr. James Joshua Pennington for their insightful remarks, as well as Dr. Yves Scherrer of the University of Helsinki for his interest in the project and ideas.

References

1. Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., Hua, X.: Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution* 6, 163–173 (2022). <https://doi.org/10.1038/s41559-021-01604-y>
2. Gonzalez, M., Aronson, B., Kellar, S., Walls, M., Greenfield, B.: Language as a Facilitator of Cultural Connection. *ab-Original* 1(2), 176–194 (2017). <https://doi.org/10.5325/aboriginal.1.2.0176>
3. Oster, R., Grier, A., Lightning, R., Mayan, M., Toth, E.: Cultural continuity, traditional Indigenous language, and diabetes in Alberta First Nations: a mixed methods study. *International Journal for Equity in Health* 13, 92 (2014). <https://doi.org/10.1186/s12939-014-0092-4>
4. Culture, Heritage and Leisure: Speaking Aboriginal and Torres Strait Islander Languages. In: 4725.0 - Aboriginal and Torres Strait Islander Wellbeing: A focus on children and youth. Australian Bureau of Statistics (2011). <https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1E6BE19175C1F8C3CA257A0600229ADC>
5. Hallett, D., Chandler, M., Lalonde, C.: Aboriginal language knowledge and youth suicide. *Cognitive Development* 22(3), 392–399 (2007). <https://doi.org/10.1016/j.cogdev.2007.02.001>
6. Whalen, D., Lewis, M., Gillson, S., McBeath, B., Alexander, B., Nyhan, K.: Health effects of Indigenous language use and revitalization: a realist review. *International Journal for Equity in Health* 21, 169 (2022). <https://doi.org/10.1186/s12939-022-01782-6>
7. Skrodzka, M., Hansen, K., Olko, J., Bilewicz, M.: The Twofold Role of a Minority Language in Historical Trauma: The Case of Lemko Minority in Poland. *Journal of Language and Social Psychology*. 39(4) 551–566 (2020). <https://doi.org/10.1177/0261927X20932629>
8. Zhang, S., Frey, B., Bansal, M.: ChrEn: Cherokee-English Machine Translation for Endangered Language Revitalization. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 577–595. Association for Computational Linguistics, Online (2020). <http://dx.doi.org/10.18653/v1/2020.emnlp-main.43>

9. Duć-Fajfer, O.: *Literatura a proces rozwoju i rewitalizacja tożsamości językowej na przykładzie literatury lemkowej* (in Polish). In: Olko, J., Wicherkiewicz, T., Borges, R. (eds.) *Integral Strategies for Language Revitalization*, 1st edn., pp. 175–200. Faculty of “Artes Liberales”, University of Warsaw, Warsaw (2016). <https://culturalcontinuity.al.uw.edu.pl/resource/integral-strategies-for-language-revitalization/>
10. Shevelov, G.: *A Historical Phonology of the Ukrainian Language* (Ukrainian translation). Vakulenko, S., Danylenko, A. (trans.), Ushkalov, L. (ed.). Naukove vydavnytstvo “AKTA”, Kharkiv (2002, original work published 1979). <http://irbis-nbuv.gov.ua/ulib/item/UKR0001641>
11. Rieger, J.: *Stanovysko i zrižnycjuvanja "rusynskŷx" dialektiv v Karpatax* (in Rusyn). In: Magosci, P. (ed.) *Najnowsze dzieje języków słowiańskich. Rusyn'skŷj jazŷk*, pp. 39–66. 2nd edn. Uniwersytet Opolski — Instytut Filologii Polskiej, Opole (2007). <https://www.unipo.sk/cjknm/hlavne-sekcie/urjk/vedecko-vyskumna-cinnost/publikacie/26405/>
12. Vaňko, J.: *Klasifikacija i holovný znakŷ Karpatskŷx Rusyn'skŷx dialektiv* (in Rusyn). In: Magosci, P. (ed.) *Najnowsze dzieje języków słowiańskich. Rusyn'skŷj jazŷk*, pp. 67–84. 2nd edn. Uniwersytet Opolski — Instytut Filologii Polskiej, Opole (2007). <https://www.unipo.sk/cjknm/hlavne-sekcie/urjk/vedecko-vyskumna-cinnost/publikacie/26405/>
13. Vaňko, J.: The Rusyn language in Slovakia: between a rock and a hard place. In: Duchêne, A. (ed.) *International Journal of the Sociology of Language*, vol. 2007, no. 183, pp. 75–96. Walter de Gruyter GmbH, Berlin (2007). <https://doi.org/10.1515/IJSL.2007.005>
14. Sopolyha, M.: *Do pytan' etničnoj identyfikacii ta sučasnyx etničnyx procesiv ukraïnciv Prŷašivščyny* (in Ukrainian). In: Skrypnyk, H. (ed.) *Ukraïnci-rusyny: etnol'inhvistyčni ta etnokul'turni procesy v istoryčnomu rozvytku*, pp. 454–487. National Academy of Sciences of Ukraine, National Association of Ukrainian Studies, Rytsky Institute of Art Studies, Folklore and Ethnology, Kyiv (2013). <http://irbis-nbuv.gov.ua/ulib/item/UKR0001502>
15. Orynycz, P.: Say It Right: AI Neural Machine Translation Empowers New Speakers to Revitalize Lemko. In: Degen, H., Ntoa, S. (eds.) *Artificial Intelligence in HCI. HCII 2022. Lecture Notes in Computer Science*, vol 13336, pp. 567–580. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-05643-7_37
16. Orynycz, P., Dobry, T., Jackson, A., Litzenberg, K.: Yes I Speak... AI neural machine translation in multi-lingual training. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC) 2021*, Paper no. 21176. National Training and Simulation Association, Orlando (2021). <https://www.xcdsystem.com/iitsec/proceedings/index.cfm?Year=2021&AbID=96953&CID=862>
17. Scherrer, Y., Rabus, A.: Neural morphosyntactic tagging for Rusyn. In: Mitkov, R., Tait, J., Boguraev, B. (eds.) *Natural Language Engineering*, vol. 25, no. 5, pp. 633–650. Cambridge University Press, Cambridge (2019). <https://doi.org/10.1017/S1351324919000287>
18. Horoszczak, J.: *Słownik lemkowski-polski, polsko-lemkowski* (in Polish). Rutenika, Warsaw (2004).
19. Pyrtej, P.: *Korotkyj slovnyk lemkiw's'kyx hovirok* (in Ukrainian). Siversiya MV, Ivano-Frankivsk (2004).
20. Duda, I.: *Lemkiw's'kyj slovnyk* (in Ukrainian). Aston, Ternopil (2011).
21. Rieger, J.: *Słownictwo i nazewnictwo lemkowe* (in Polish). Wydawnictwo naukowe Semper, Warsaw (1995).
22. Fontański, H., Chomiak, M.: *Gramatyka języka lemkowego* (in Polish). Wydawnictwo Naukowe „Śląsk”, Katowice (2000).

23. Pyrtej, P.: *Dialekt łemkowski. Fonetyka i morfologia* (in Polish). Hojsak, W. (ed.). Zjednoczenie Łemków, Gorlice (2013).
24. Post, M.: A Call for Clarity in Reporting BLEU Scores. In: Proceedings of the Third Conference on Machine Translation (WMT), vol. 1, pp. 186–191. Association for Computational Linguistics, Brussels (2018). <https://doi.org/10.48550/arXiv.1804.08771>
25. Papineni, K., Roukos, S., Ward, T., Wei-Jing, Z.: BLEU: a Method for Automatic Evaluation of Machine Translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 02), pp. 311–318. Association for Computational Linguistics, Philadelphia (2002). <https://doi.org/10.3115/1073083.1073135>
26. Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J.: A Study of Translation Edit Rate with Targeted Human Annotation. In: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers, pp. 223–231. Association for Machine Translation in the Americas, Cambridge (2006). <https://aclanthology.org/2006.amta-papers.25>
27. Popović, M.: CHRF: character n -gram F-score for automatic MT evaluation. In: Proceedings of the Tenth Workshop on Statistical Machine Translation, pp. 392–395. Association for Computational Linguistics, Lisbon (2015). <http://dx.doi.org/10.18653/v1/W15-3049>