Yes I Speak… AI Neural Machine Translation in Multi-Lingual Training

Péter Orynycz, Tom Dobry, Angela Jackson, and Kourtni Litzenberg
Antech Systems, Inc.
Lexington Park, Maryland, United States of America
p@orynycz.com, {tom.dobry, angie.jackson, kourtni.litzenberg}@antechsystems.com

ABSTRACT

Rapidly sharable and jointly usable training across coalition partners needs to linguistically and culturally adapt (that is, *localize*) to the languages of non-Anglosphere allies representing frontline actors with limited or potentially no English proficiency. Traditionally, localization has been a time and labor-intensive process, with an expert needing over two months to translate a mid-sized book. Meanwhile, training exercises are charged with responding to total development timelines of days and weeks, not months and years, in order to respond to evolving realities of the operational world. In this frame, traditional localization becomes a major bottleneck for coalition endeavors. In rapidly unfolding situations, allies simply cannot wait months and years on training needed tonight or to address next week's mission in languages readily understood on the front line—that is, to realize the long-aspired dream of locally tailored training to address boots-on-the-ground needs.

Enter the emergent artificial intelligence technology of *neural machine translation*, which can do in mere minutes that which takes seasoned organic linguists an hour, making coalition-wide, multilanguage deployment in days and weeks feasible, with ever improving complexity accounted for. This is made possible by machine learning, that is, training artificial recurrent neural networks to translate from one natural language to another.

We built artificial-intelligence based engines, timed their translation of North Atlantic Treaty Organization (NATO) training materials, and measured their accuracy using the bilingual understudy evaluation (BLEU) metric. Our engine translated from Russian 1,169.51% faster and 58.37% more accurately than our professional human linguist used as a control. Our Polish neural engine was 17.29% more accurate and 488.45% faster than human. Our Lemko translation engines are the world's first and scored a decent BLEU of 14.57. Meanwhile, we did the above on an inexpensive laptop computer in an air-gapped, access-controlled environment cut off from the outside world.

ABOUT THE AUTHORS

Péter Orynycz (Cyrillic: **Петро Оринич**) (ORCID iD https://orcid.org/0000-0003-3094-9156) is a senior software developer at Antech Systems, Inc., computer programmer, computational linguist, localization and natural language engineer, Slavist, and Silicon Valley consultant. His research currently focuses on artificial intelligence, neural machine translation, and hybrid systems for endangered, indigenous language revitalization. He received his degree in Russian at the Institute of East Slavic Philology of Jagiellonian University in Cracow, Poland, where he worked for Google amid its 2016 neural machine translation artificial intelligence breakthrough. His engines were recently mentioned in the Cambridge University Press journal *Natural Language Engineering* (Volume 25, Issue 5, page 634). Mr. Orynycz also has two decades of transatlantic experience as a linguist specializing in Russian, Polish, Ukrainian, Rusyn and Lemko for top language service providers, national defense, heavy industry, Raytheon, Amazon, Siemens, Mercedes-Benz, Daimler, investigators, philanthropists, and scientists.

Angela Jackson is a software developer at Antech Systems, Inc., collaborating with the ePerformance Solutions team in creating interactive courseware. She has supported various Department of Defense programs for over 20 years, with ten years of that experience focused on designing and developing systems training courses. Ms. Jackson is currently working towards a Master of Science degree in Information Systems with a specialization in web development.

Kourtni Litzenberg is a software developer for the Antech Systems, Inc. DEFTECH team with a software testing background. She has an Associate of Science degree in Software Engineering and is CompTIA Security+ certified. Her testing background has been used to assist in testing other projects such as Joint Incident Site Communications Capability (JISCC) and Civil Support Team (CST). During her time with Antech, she has done research mostly on microservices and automated testing.

Yes I Speak… AI Neural Machine Translation in Multi-Lingual Training

Péter Orynycz, Tom Dobry, Angela Jackson, and Kourtni Litzenberg
Antech Systems, Inc.
Lexington Park, Maryland, United States of America
p@orynycz.com, {tom.dobry, angie.jackson, kourtni.litzenberg}@antechsystems.com

INTRODUCTION

The Problem

Online translation services work until one needs to translate in secret, or from a language not among the top 2% in terms of resources. Using even the most secure online or cloud translation services means sharing information with a third party, which violates most commercial non-disclosure agreements, not to mention defense-industry requirements. Next, the leading online translation platform works for only 109 languages (Google, 2021), which is less than 2% of the 7,139 spoken in the world today (Eberhard, Simons, & Fennig, 2021). Faced with the inherent third-party risk of cloud service providers, many turn to calling in vetted human linguists, who can at best handle a book a month, and at worst pose lifelong unauthorized-disclosure threats. To give multilingual enterprises and organizations more options, we set out to solve the problem of employing artificial intelligence to quickly, accurately, and covertly translate materials from high-resource, medium-resource, and low-resource languages on air-gapped, inexpensive, mid-range laptop computers disconnected from the internet and outside world.

Solutions So Far

While the groundwork for machine translation was laid in al-Kindī's Baghdad over a millennium ago (DuPont, 2018; see also al-Kindī, 2002), almost all the spectacular, visible progress has taken place in Silicon Valley over the last five years. The main breakthrough came at Google (Lewis-Kraus, 2016), and Facebook has now joined the neural machine translation club (Ott, et al., 2019). We built upon the work of their *FAIRseq* engine, for which Sławomir Dadas has made available an excellent Polish-English model (Dadas, 2019). We built on Mr. Dadas' work to craft hybrid neural/rule-based/dictionary-based engines that translate from Lemko to English, and vice versa. The idea of applying *transfer learning* for Rusyn natural language processing (NLP) had been discussed with our esteemed colleagues Yves Scherrer and Achim Rabus, who were first to publish results in a peer-reviewed journal, and kindly took the opportunity to mention Péter Orynycz's hybrid neural/rule-based Lemko machine translation engine (Scherrer & Rabus, Neural morphosyntactic tagging for Rusyn, 2019, p. 634), which has been operational and freely available for public use at the web address `www.lemkotran.com` since March 2019, with its transliteration natural-language processing module made available for public use there in September 2017.

The Wider System

High, medium, and low-resource languages
Language pairs are classified in scientific literature as being *high-resource*, *medium-resource*, and *low-resource*, depending on the amount of technologies and data sets available relative to their international importance (Cieri, Maxwell, Strassel, & Tracey, 2016, p. 4545). High resource pairs include Czech-English (Kocmi, 2020, p. 171), Russian-English, German-English (Ng, et al., 2019, p. 314) and Chinese-English (Kocmi & Bojar, 2019, pp. 234–235). Polish-English is a medium-resource pair (Jónsson, Símonarson, Snæbjarnarson, Steingrímsson, & Loftsson, 2020, p. 2). Low-resource pairs include Gujarati-English, Kazakh-English (Kocmi & Bojar, p. 234), Inuktitut-English (Kocmi, p. 171), and Lemko-English (Scherrer & Rabus, 2019, p. 85). Since training artificial intelligence language models requires vast amounts of bilingual data, higher resource languages generally enjoy the availability of neural machine translation engines. Meanwhile, due to a lack of machine-learning training data, neural engines are rarer for lower resource languages, which are often better-served by previous-generation *statistical machine translation* (SMT) engines.

High-resource language under experiment: Russian
Russian is spoken as a first language by more than 168 million, and as an additional language by another 114 million (Maximova, Noyanzina, Omelchenko, & Maximova, 2018, p. 2). Automating its translation into English has been the holy grail of postwar machine translation efforts. As one of the official languages of the United Nations, enormous amounts of bilingual Russian-English text are available under a liberal license (Ziemski, Junczys-Dowmunt, & Pouliquen, 2016, p. 3530).

Medium-resource language under experiment: Polish
Polish is a West Slavic language spoken by about 38 million within today's Poland, a number expected to fall due in part to the pandemic ongoing at the time of publication (Associated Press, 2021). A further 10 million speak Polish to some degree beyond the nation's borders (Jassem, 2003, p. 103). As one of the official languages of the European Union, large amounts of bilingual text are available to train artificial intelligence translation models, including 22,630 European Parliament documents (Hajlaoui, Kolovratnik, Vaeyrynen, Steinberger, & Varga, 2014, p. 3165).

Low-resource language under experiment: Lemko
Lemko is a low-resource language (Scherrer & Rabus, 2019, p. 85) that meets traditional criteria for classification as East Slavic. For example, Lemko exhibits East Slavic pleophony, that is, the outcome of Proto-Slavic "*ToRT*" sequences is *ToRoT* (Fortson IV, 2004, pp. 371-372), as in Lemko *horodyty* 'to fence, to enclose' (Horoszczak, 2004, p. 45), as well as in standard Ukrainian *horodyty*, Rusyn *horodyty*, and Russian *gorodit'* (Kerča, 2007, p. 176). Meanwhile, compare Polish (a West Slavic language) with *-ro-* in *grodzić*, but Croatian (a South Slavic language) with *-ra-* in *graditi,* 'build'. Further afield, there is English with *-ar-* in *yard* and *garden*, Avestan (Old Iranian) with *-ərə-* in *gərəðō* 'cave', and Sanskrit (Old Indic) with *-ṛ-* in *gṛhás* 'home' (Vasmer, p. 1443).

While the exact classification of Lemko and its status relative to Standard Ukrainian and codified Rusyn is a matter of controversy (Rabus & Scherrer, 2017), our Lemko-English engine scoring so high without recourse to resources of Standard Ukrainian or Rusyn as codified in Slovakia could lend support to the conclusion of Watral (2015) that Lemko is a full-fledged language unto itself, and not a dialect of any other tongue. Buoyed by rising objective quality scores, we decided to prioritize Polish transfer learning due to its immediate return on investment in terms of Lemko translation accuracy, our highest value. It is possible that quality scores were boosted by interference from observed hybrid language whereby Lemko grammatical endings are retrofitted onto Standard Polish words (Watral, 2016, p. 242).

Poland's census bureau tallied 6,279 as speaking Lemko at home in 2011, up from 5,605 in 2002 *(Departament Wyznań Religijnych oraz Mniejszości Narodowych i Etnicznych*, 2013, p. 7), with a new count underway at the time of publication. How many of the 24,539 residents of Poland counted as speaking Ukrainian at home or 626 speaking "Ruthenian" (*język ruski*) with other household members in 2011 *(Departament Wyznań Religijnych oraz Mniejszości Narodowych i Etnicznych*, 2013, p. 7) might be Lemko speakers is beyond the scope of this paper. Ukraine's State Statistics Service has counted 672 Lemkos within its borders (*Deržavna služba statystyky Ukraïny*, 2001). On the United Nation's language endangerment scale of 0 to 5, with 0 being extinct and 5 "safe" (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003, pp. 7-8), Lemko would be approaching 2, that is, *seriously endangered*: natural intergenerational language transmission is increasingly absent, and younger speakers are increasingly non-existent (Duć-Fajfer, 2016, p. 178). There are shoots of green, however, with laws that protect and promote minority language use in education, broadcasting, publishing, road signage, and science increasingly being taken advantage of (Duć-Fajfer, 2016, pp. 178-179).

The resource situation is improving as well. Péter Orynycz has compiled and aligned a bilingual Lemko-English corpus comprising 68,599 source words together with his translations into English (the only existent parallel text of which we are aware). The corpus was put together using interviews conducted in Lemko by the John and Helen Timo Foundation of the United States, who commissioned Mr. Orynycz to have them transcribed and translate them, as well as permitted him to use the work in his scientific research and development. He is also amassing a monolingual Lemko corpus of over a million words. While the complex sociolinguistic relationships between Lemko, Rusyn, standard Ukrainian, and Slovak language communities are a matter beyond the scope of this paper, it was Polish resources (specifically, Polish neural models) that were instrumental to Mr. Orynycz's hybrid Lemko engines.

HYPOTHESES AND PREDICTIONS

Translation Speed

Hypothesis: air gapped artificial-intelligence machine translation now as fast as humans
We hypothesized that neural machine translation engines running offline on mid-range laptops are now of comparable speed to human translators. This was based on observations during engine development that neural machine translation seemed to take from several seconds to under a minute to translate a sentence on a mid-range laptop, which is comparable to the human speeds observed by Péter Orynycz in his experience in the localization industry.

Prediction: machine translation engines will process more words per hour than human translators
Based on our hypothesis that air-gapped neural machine translation engines running offline on mid-range laptops would be as fast as humans, we predicted that their speed would exceed that of human linguists, and that they would translate more words per second than our human control subject.

Translation Accuracy

Hypothesis: artificial-intelligence machine translation engines are now almost as accurate as human translators
We hypothesized that neural machine translation engines were now almost as accurate as human translators. This was based on Péter Orynycz's professional observation as a translation quality control specialist that commercial neural machine translation cloud services had not only dramatically improved, they were producing results often indistinguishable from those of human linguists.

Prediction: artificial-intelligence machine translation engines will achieve at least 75% the BLEU quality score of professional human translators
While we knew neural machine translation engines could be superior to bilingual amateurs trying their hand at translating for the first time, we did not believe our engines would beat experienced, professional linguists in a head-to-head competition. Fortunately, our doubt could be put to the test. The bilingual evaluation understudy (BLEU) algorithm is the most dominant metric for machine translation research, being language-independent, cheap and easy-to-compute, as well as reasonably correlated with human judgements (Post, 2018). We predicted our neural engines would score 75% the quality points scored by a human linguist. For example, if a human linguist scored 40, the neural machine translation would score 30. Meanwhile, we predicted our hybrid Lemko-English engine would achieve a cumulative BLEU score of 15.

Translation Security

Hypothesis: artificial-intelligence machine translation can be performed offline on laptops in high security field settings
We hypothesized that neural machine translation could be performed offline on air-gapped, portable equipment completely cut off from the outside world. This was based on the observation that all of the components of our solution made no calls to the internet once dependencies had been installed. An implicit assumption is that air-gapped translation systems with *Airplane Mode* enabled cannot be remotely monitored or hacked. Another assumption is that not only have operators been appropriately vetted, they have taken appropriate precautions against external and insider threats. Another implicit assumption is that it is easier to safeguard just one mobile workstation for several hours than prevent human linguists making an average of USD 25.01 per hour (Bureau of Labor Statistics, United States Department of Labor, 2021) from making unauthorized disclosures over the course of a lifetime, especially in light of reports of linguists being arrested on suspicion of leaking secrets (Department of Justice Office of Public Affairs, 2009, 2018, 2020).

Prediction: artificial-intelligence machine translation will succeed on an air gapped Lenovo Legion Y730-17ICH laptop computer running offline in Airplane Mode

We predicted our translation system would not malfunction and would complete its tasks when physically separated and disconnected from any and all networks or devices by activating the Airplane Mode feature of Windows 10 Pro on a Lenovo Legion Y730-17ICH laptop computer (Type 81HG).

Hybrid Rule/Dictionary-based and Neural Lemko-English Engine

Hypothesis: hybrid dictionary/rule-based engines improve machine translation accuracy

We hypothesized that our Polish-Lemko rule-based machine translation (RBMT) engine, Polish-Lemko dictionary-based machine translation (DBMT) engine, Lemko-Polish DBMT engine run in reverse, and neural Polish-English engine could be synergistically coupled into a hybrid engine that achieves higher quality scores with each additional part. This hypothesis was based on the author's observations working as a professional Lemko-English translator that correspondences between Lemko and Polish were frequent enough to make a hybrid engine a viable proposition.

Prediction: each sub-engine added to our hybrid Lemko-English engine will increase BLEU by 5 points

We predicted that for each rule-based or dictionary-based Lemko-Polish sub-engine we added to our hybrid Lemko-English engine, the overall BLEU accuracy score would increase 5 points.

Introduction to Methods and Justification

We pitted man against machine by giving both a mid-range, air-gapped laptop running our custom computer-assisted translation program (detailed below) while offline in Windows Airplane Mode. We recorded speed and accuracy at translating from Russian to English (a high-resource language pair), Polish to English (a medium-resource pair), and Lemko to English (a low-resource pair). To express speed, we used the words per hour metric because it is the mainstay of localization project managers, as well as used in scientific literature (Macken, Prou, & Tezcan, 2020, p. 4). To measure accuracy, we used the BLEU metric because it is the most widespread one in the field of research and development (Post, 2018).

Principal Results in Brief

Not only were we able to apply the breakthrough technology of neural machine translation to use artificial intelligence on an air gapped, offline laptop in Airplane Mode to translate a high resource language (Russian) over 10 times faster than our human linguist control subject, our machine's quality score was over 58 percent "better than human". Moreover, we are the first team in the world to publish results for Lemko machine translation engines in a scientific journal.

MATERIALS AND METHODS

Introduction

To test our predictions, we built a number of artificial-intelligence and hybrid translation engines, calculated their speed and accuracy on an air-gapped laptop in Windows Airplane Mode, and did the same with a professional linguist in order to make our experiment a controlled one.

Lab Set Up

Hardware

We used a Lenovo Legion Y730-17ICH laptop computer (Type 81HG) running Windows 10 Pro (64-bit). The model has been discontinued and sells for about USD 850, pre-owned, as of the time of publication.

Operating system

The virtualized operating system used for the experiment was the Linux Subsystem for Windows, and to be exact, Ubuntu 18.04 LTS installed via the *Microsoft Store* digital distribution platform.

Dependencies

Python 3.8 was installed using the command `sudo apt install python3.8`.

The command `sudo python3.8 -m pip install -upgrade` was used to install major dependencies, including `bleu`, `fastBPE`, `hydra-core`, `python-dev-tools`, `PyYAML`, `omegaconf`, `pip`, `pytz`, `nltk`, `setuptools`, `sacremoses`, `subword-nmt`, `torch`, and `torchvision`.

Toolkits

We installed the Facebook AI Research Sequence-to-Sequence Toolkit by running the following commands:

```
sudo git clone https://github.com/pytorch/fairseq
cd fairseq
sudo python3.8 -m pip install --upgrade --ignore-installed PyYAML --editable ./
```

Documentation and technical support are available at `https://github.com/pytorch/fairseq`

Neural machine translation models

For our neural Polish-English and hybrid Lemko-English engines, we used Sławomir Dadas's Polish-English convolution model, available and documented at his Polish Natural Language Processing (NLP) Resources repository (Dadas, 2019).
Model: `https://github.com/sdadas/polish-nlp-resources/releases/download/nmt-models-conv/polish-english-conv.zip`
Documentation: `https://github.com/sdadas/polish-nlp-resources#machine-translation-models`

For our Russian-English engine, we leveraged the Facebook AI Research Sequence-to-Sequence (FAIRseq) Russian to English pretrained single transformer model without finetuning, which was submitted to the 2019 Fourth Conference on Machine Translation (WMT19).
Model: `https://dl.fbaipublicfiles.com/fairseq/models/wmt19.ru-en.ffn8192.tar.gz`
Documentation: `https://github.com/pytorch/fairseq/tree/master/examples/wmt19`

Electronic dictionaries

Our professional linguist was permitted offline access to the electronic versions of the *New Kościuszko Foundation American English to Polish Dictionary* (USD 12.99), as well as the *Oxford Russian Dictionary* (USD 19.99). Both were purchased via the Microsoft Store. Jarosław Horoszczak's Lemko-Polish and Polish-Lemko dictionary (2004) was also made available to our linguist for offline use.

Experiment control

The experiment was controlled by sitting a professional human linguist at the air gapped laptop while it was in Airplane Mode, with the aforementioned electronic dictionaries available on the machine. The linguist would press the enter key, at which point the timer would start and the source sentence to be translated would display. The linguist was permitted to type his translation in Microsoft Word (to take advantage of its spell check feature and other word processing aids) and then paste it into our custom computer-assisted translation program. After pressing the enter key again, the human translation was submitted and the timer stopped. The human linguist's speed in terms of words per hour and accuracy in terms of BLEU score were calculated for each sentence translated.

Péter Orynycz, who has two decades of experience as a Russian and Polish linguist, a Polish university degree in Russian, and over 5 years' experience as a professional Lemko-English translator, served as the control subject. He performed back translations of the Russian and Polish materials listed below, as well as English retranslations from Lemko.

Experiment material: reference translations
The Russian-English and Polish-English text for the experiment was obtained from educational materials shared with the public and translated from English into Russian and Polish by the *NATO Review* publication of the North Atlantic Treaty Organization (NATO). To quote, "Reproduction of parts, excerpts or articles of the NATO Review is authorized for non-commercial purposes, pursuant to the following condition: the source, NATO Review, must be acknowledged." As is standard practice (Post, 2018), the corpus data was cleaned and normalized by lowercasing text and tokenizing it. Care was taken to ensure the source text and target translations were aligned at the sentence level.

For this experiment, we used a lecture delivered by Dr. Jamie Shea, then NATO's Deputy Assistant Secretary General for Emerging Security Challenges. Its title is *What Can We Learn Today from the 'Three Wise Men'?* The English original text of Dr. Shea's lecture and its NATO-commissioned translations into Russian and Polish were retrieved from the following uniform resource locators:
English original: `https://www.nato.int/docu/review/articles/2016/12/05/what-can-we-learn-today-from-the-three-wise-men/index.html`
Russian translation: `https://www.nato.int/docu/review/ru/articles/2016/12/05/chemu-my-moyoem-nauchit-sya-segodnya-u-treh-mudretsov/index.html`
Polish translation: `https://www.nato.int/docu/review/pl/articles/2016/12/05/czego-mozemy-nauczyc-sie-dzisiaj-od-trzech-medrcow/index.html`

The Lemko-English material for the experiment comprised in-person interviews recorded by the John & Helen Timo Foundation of the United States, who had hired Péter Orynycz to have the interviews transcribed and translate them into English. The foundation later kindly donated the resulting bilingual corpora to scientific research and development. To protect the privacy of those discussed in the interviews, and out of respect for the European Union General Data Protection Regulation (GDPR), the materials have not been made publicly available. Care is taken to redact any personally identifiable information (PII) and personal health information (PHI) before sharing samples.

Method for scoring translation accuracy: BLEU
The bilingual understudy evaluation (BLEU) metric was used to measure similarity to the reference translation, and thus, however imperfectly, accuracy. Though the BLEU score is not a perfect measure of accuracy or quality, it is the one most widely used in the industry (Post, 2018). The Python module was obtained from the Python `bleu` package, documented at the following uniform resource locator: `https://pypi.org/project/bleu/`

We made sure that when given the reference sentence strings "`it is a white cat .`" and "`wow , this dog is huge .`" together with the candidate hypotheses "`it is a white kitten .`" and "`wowww , the dog is huge !`", our system calculated a cumulative BLEU score of 34.99, in line with the documentation for the Python `bleu` package.

Method for normalizing and cleaning text
All text was lowercased and a space was added before and after all punctuation marks so that the system would not assume, for example, that "`Cat`" and "`cat.`" were different words. Thus, "`It is a white cat.`" would be normalized to "`it is a white cat .`" Multiple spaces and other whitespace sequences were replaced with a single space using the Python `split()` and `join()` methods.

Method for counting words
The number of words per sentence was determined by splitting the normalized text string into an array using a space as a delimiter, and then counting the items in that array.

Method for measuring translation speed
The moment the human translator pressed the enter key to start translating a sentence, we called the Python `time.time()` method to get the number of seconds since midnight Coordinated Universal Time (UTC) on January 1, 1970 as a floating point number, commonly referred to as *Unix time*, and used that number as the human start time. We also obtained the Unix time when our machine translation engines took up a sentence for translation.

The moment a human linguist pressed the enter key to submit their translation of a sentence or a machine returned its translation of a sentence was used as the end time. By subtracting the start time from the end time, we arrived at the total number of seconds the translation of a sentence took.

We calculated words per hour by dividing the difference between the Unix end and start times by the wordcount as calculated above, and multiplying that quotient by 3600 (which is 60×60, that is, the number of seconds in an hour):

$$Speed = \frac{(Tranlation\ end\ time - Translation\ start\ time)}{Total\ number\ of\ words\ translated} \times 3600 \qquad (1)$$

Method for physically isolating and air gapping equipment
The equipment used in the experiment was cut off from the outside world not only by means of its physical isolation, but also by employing the Airplane Mode feature of Microsoft Windows 10 Pro, which, per its documentation, turns off all wireless communication on the machine, including IEEE 802.11b Direct Sequence wireless networking, cellular, Bluetooth, Global Positioning System, and Near Field Communication.

RESULTS

Translation Speed: the Higher-Resource the Language Pair, the Faster the Engine

Machine bested man in terms of translation speed for the high-resource pair of Russian-English and medium-resource pair of Polish-English, in line with our hypothesis that neural machine translation was faster than humans and prediction that our neural engines would translate more words per hour. When translating from Russian, our engine averaged over 6,456 words per hour, which was 1,170% faster than our human linguist. When translating from Polish, our neural engine was 488% faster than our human translator, averaging 3,768 words per hour. For the low-resource pair of Lemko-English, our hybrid neural and dictionary/rule-based engine managed 707 words per hour, nearly tying our human linguist, who was 13% faster at 798 words per hour. Removing the weight of the dictionary-based component of the hybrid engine nearly quadrupled speed to 3,137 words per hour, which is 293% faster than human, at the cost of a 13% drop in accuracy.

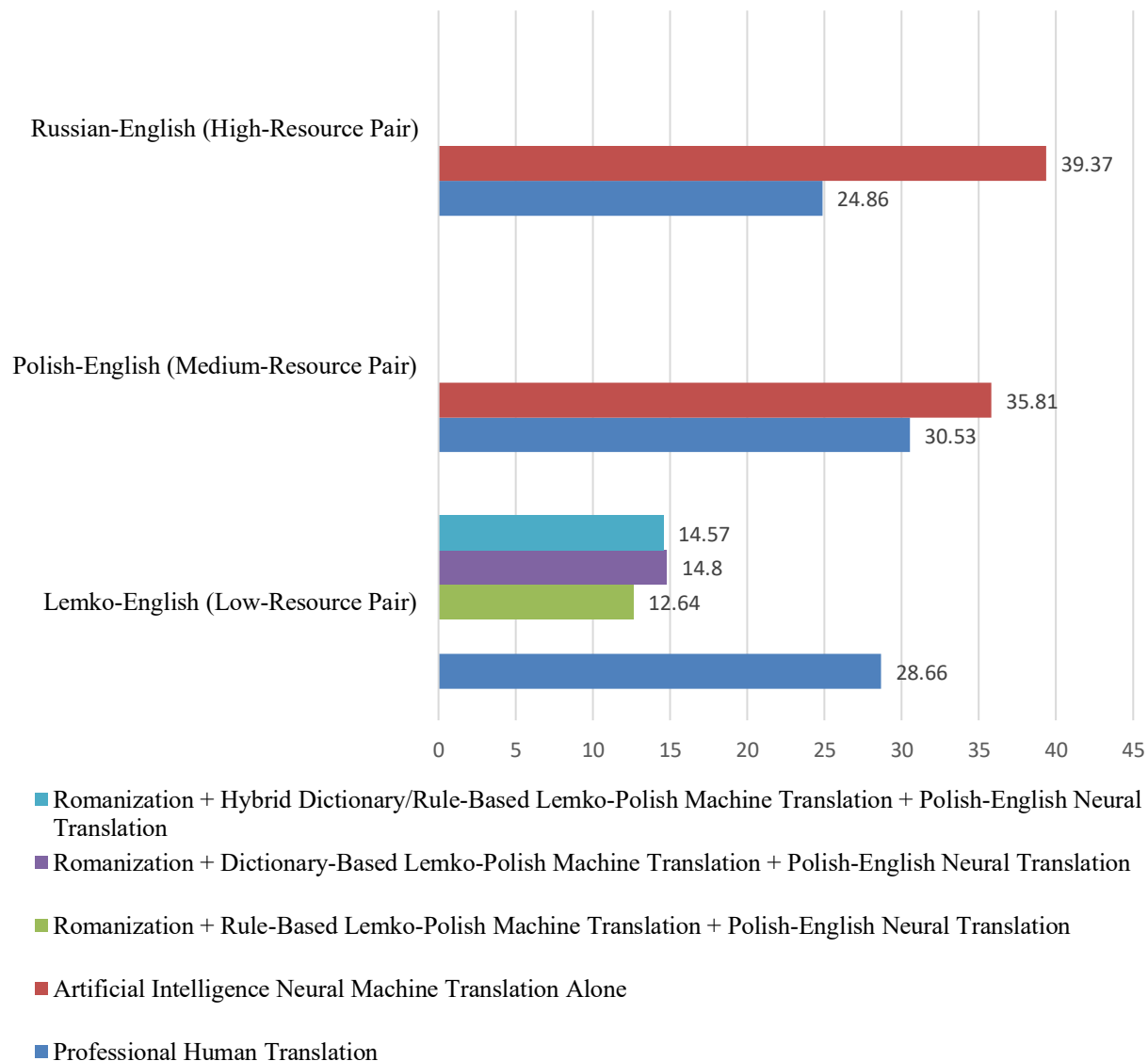
Human versus Machine Translation Speed, Words per Hour
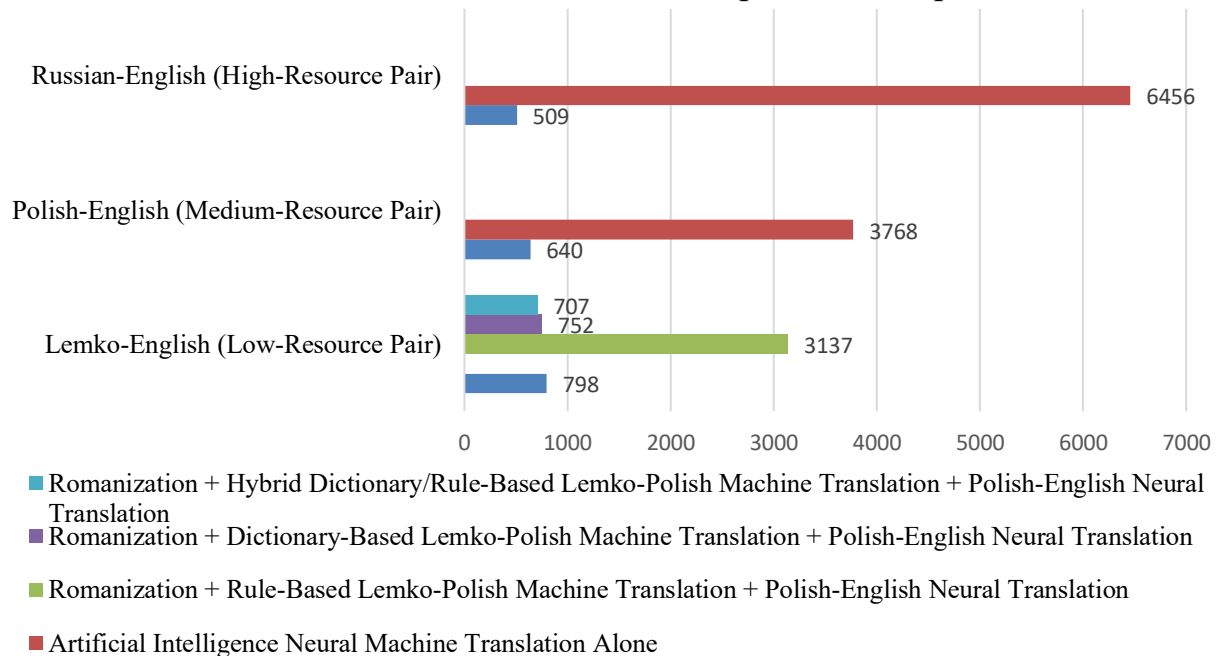
Figure 1. Professional Human versus Machine Translation Speed in Words per Hour on an Air-Gapped Mid-Range Laptop in Airplane Mode, Russian-English (High Resource Pair) versus Polish-English (Medium-Resource Pair) versus Lemko-English (Low-Resource Pair)

Translation Accuracy: the Higher-Resource the Language Pair, the More Accurate the Engine

The translation accuracy of our artificial intelligence engines surpassed that of professional linguists. This went above and beyond our hypothesis that air-gapped neural machine translation was now only slightly less accurate than human translators. Our Russian-English artificial-intelligence engine achieved 158% the accuracy of our human translator, exceeding the 75% we predicted. Our Polish-English neural engine scored 117% the accuracy of our human linguist, exceeding our 75% expectation. Our hybrid Lemko-English engine achieved a BLEU score of 14.57 (51% that of our professional translator), in line with our prediction of 15, when rounded up. Dropping our rule-based sub-engine resulted in a 2% gain in accuracy and 6% increase in speed. Dropping the dictionary-based sub-engine resulted in a 13% drop in accuracy, but a 344% increase in speed. To summarize, our medium-to-high resource language, artificial-intelligence engines were significantly more accurate than our human linguist, while our low-resource language, hybrid engines were about half as accurate as our human linguist.

Human versus Machine Translation Accuracy, BLEU Score

Legend:
- ■ Romanization + Hybrid Dictionary/Rule-Based Lemko-Polish Machine Translation + Polish-English Neural Translation
- ■ Romanization + Dictionary-Based Lemko-Polish Machine Translation + Polish-English Neural Translation
- ■ Romanization + Rule-Based Lemko-Polish Machine Translation + Polish-English Neural Translation
- ■ Artificial Intelligence Neural Machine Translation Alone
- ■ Professional Human Translation

Figure 2. Professional Human versus Machine BLEU Translation Quality Score on an Air-Gapped Mid-Range Laptop in Airplane Mode, Russian-English (High-Resource Pair) versus Polish-English (Medium-Resource Pair) versus Lemko-English (Low-Resource Pair)

Translation Security

In line with our hypothesis that a neural machine translation solution could be engineered to run on an air-gapped laptop, our experiment succeeded in that regard. In keeping with our prediction, our experiment worked with Windows Airplane Mode enabled, and no errors were caused by operating while cut off from the outside world.

Rule-Based Machine Translation between Lemko and Polish

Our hypothesis that affinity between Lemko and Polish was strong enough that Lemko would be translatable into Polish using rule-based and dictionary-based substitution was proven correct by the impressive performance of our Lemko-English neural/rule-based hybrid engine. Our hypothesis that combining a rule-based sub-engine with a dictionary-based one would result in a more accurate hybrid engine is not supported by our data at this time. Adding a dictionary-based module to a rule-based one increased engine BLEU by 2.16 points, less than our prediction of 5.

DISCUSSION

A new era

We proved that not only is it possible to task artificial intelligence with the knowledge work of translation from high, medium, and low resource languages in an access-controlled environment, neural machine translation can do the work faster, more securely, and in many cases, better. Not only did our results support our hypotheses, the performance of our neural engines surpassed our predictions. A new era of near real time machine translation acting independently or in partnership with humans is here.

Speed

Our engine translated from Russian at a rate of 6,456 words per hour. To put that into context, we consulted with expert Marc Hackel, a Washington, D.C. defense industry linguist and Russian-English translator with decades of experience, who told us that "a kind of rule of thumb is that a very accomplished translator should be able to do at least 8 pages (that is, 8 pages of 500 words for a total of 4,000 words) over an 8-hour workday, assuming there are no obstacles like acronyms and other things like that. The average for many is actually 250 words per hour, not 500." So, neural engines can do in under an hour that which takes humans days.

Accuracy

Our artificial intelligence engines achieved higher BLEU scores than our professional human linguist. By that metric, our machines are "better" at translating from Russian and Polish than people.[1] Since the implication that artificial intelligence machine translation can be over 50% more accurate than seasoned linguists is revolutionary, this experiment needs to be repeated on even more human linguists and corpora in order to rule out flukes. We used exceptionally clean, challenging texts with flowery language, which machine translation engines traditionally choke on and humans excel at. While we strove for an even playing field, we expected any advantage would be on the human side. Please contact Péter Orynycz at the address given above for access to our raw data and results.

---

[1] Historically, some communities push back on using BLEU to compare human to machine translation, yet no other system is as widely accepted or available with broad, peer reviewed validation in use. In fact, BLEU score inventors Papineni, Roukos, Ward, and Zhu foreshadowed this very point of tension in work sponsored by the United States Department of Defense (funded by the Defense Advanced Research Projects Agency [DARPA] and monitored by the Space and Naval Warfare Systems Command [SPAWAR]) as part of their seminal publication, writing: "Furthermore, it [the metric] must distinguish between two human translations of differing quality. This latter requirement ensures the continued validity of the metric as MT [machine translation] approaches human translation quality." Breaking the taboo from the outset, they then proceeded to calculate BLEU scores for "Human-1", native in neither Chinese nor English, and "Human-2", a native speaker of English, and charted how their BLEU scores tracked closely to those given by human judges (Papineni, Roukos, Ward, & Zhu, 2002).

Next steps

We used older, legacy equipment. Newer equipment with faster, next-generation graphical processing units could result in a dramatic improvement in translation speed. Our codebase should be optimized to maximize use of existing resources like graphics processing units (GPUs). We plan to convert our dictionary-based machine translation module into a test suite for use in test-driven development (TDD) of our rule-based machine translation (RBMT) module, which could be used to develop parallel texts for training purely neural Lemko-English and English-Lemko artificial intelligence neural machine translation engines. More research is needed to identify points of diminishing returns. Péter Orynycz plans to apply his hybrid neural and rule-based systems to develop translation engines for Rusyn and Ukrainian dialects indigenous to today's Slovakia and Ukraine.

In closing

We are at the dawn of a new transformative era: we proved artificial intelligence can perform knowledge work as well as humans, or in a widening set of cases, over 50% better, and in a fraction of the time and with almost none of the security risk. A few hundred dollars' worth of equipment that fits in a backpack is all one needs to always have a better-than-human, silicon-based field linguist sidekick who never spills secrets or tires. The genie is out of the bottle, and may grant our wish of revitalizing endangered languages, if not the dream of raising extinct ones from the dead. The language mass extinction event we are in the middle of might grind to a halt, and even reverse. We should be careful what we wish for—worlds insulated for eons by their encryption in expensive-to-translate languages are set to collide. A change for the better, we hope. *Proščaj*, language barrier. Hello, new world.

ACKNOWLEDGEMENTS

REFERENCES

al-Kindī, Y. i. (2002). al-Kindi's Edited Treatise. In M. I. AL-Suwaiyel, I. A. Kadi, & M. al-Bawab (Eds.), *al-Kindi's Treatise on Cryptanalysis (Vol. 1)* (S. M. al-Asaad, Trans., Vol. 1, pp. 117-204). Damascus, Syria: KFCRIS & KACST. (Original work published ca. 850).

Associated Press. (2021, January 26). *Poland's population rapidly shrinking under pandemic.* Pobrano June 19, 2021 z lokalizacji AP NEWS: https://apnews.com/article/pandemics-demographics-coronavirus-pandemic-birth-rates-covid-19-pandemic-5895d554be280b0ade9068c75872976e

Bureau of Labor Statistics, United States Department of Labor. (2021). *Occupational Outlook Handbook, Interpreters and Translators.* Washington, DC. Retrieved June 1, 2021, from https://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm

Cieri, C., Maxwell, M., Strassel, S., & Tracey, J. (2016). Selection Criteria for Low Resource Language Programs. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp.

4543–4549). Portorož, Slovenia: European Language Resources Association (ELRA). Retrieved June 27, 2021, from https://www.aclweb.org/anthology/L16-1720

Dadas, S. (2019). A repository of Polish NLP resources. Retrieved May 26, 2021, from https://github.com/sdadas/polish-nlp-resources/

Departament Wyznań Religijnych oraz Mniejszości Narodowych i Etnicznych. (2013). *IV Raport dotyczący sytuacji mniejszości narodowych i etnicznych oraz języka regionalnego w Rzeczypospolitej Polskiej - 2013.* Warszawa, Poland: Ministerstwo Spraw Wewnętrznych i Administracji. Retrieved June 13, 2021, from http://mniejszosci.narodowe.mswia.gov.pl/download/86/14637/TekstIVRaportu.pdf

Department of Justice Office of Public Affairs. (2009, December 17). *Former FBI Contract Linguist Pleads Guilty to Leaking Classified Information to Blogger.* Retrieved June 9, 2021, from United Stastes Department of Justice: https://www.justice.gov/opa/pr/former-fbi-contract-linguist-pleads-guilty-leaking-classified-information-blogger

Department of Justice Office of Public Affairs. (2018, August 23). *Federal Government Contractor Sentenced for Removing and Transmitting Classified Materials to a News Outlet.* Retrieved June 9, 2021, from United States Department of Justice: https://www.justice.gov/opa/pr/federal-government-contractor-sentenced-removing-and-transmitting-classified-materials-news

Department of Justice Office of Public Affairs. (2020, August 17). *Former CIA Officer Arrested and Charged with Espionage.* Retrieved June 9, 2021, from United States Department of Justice: https://www.justice.gov/opa/pr/former-cia-officer-arrested-and-charged-espionage

Deržavna služba statystyky Ukraïny. (2001). *Čysel'nist' osib okremyx etnohrafičnyx hrup ukrainskoho etnosu ta ïx ridna mova.* Pobrano August 26, 2021 z lokalizacji Vseukraïns'kyj perepys naselennja 2001: http://2001.ukrcensus.gov.ua/results/nationality_population/nationality_popul2/select_5/?botton=cens_db&box=5.5W&k_t=00&p=0&rz=1_1&rz_b=2_1&n_page=1

Duć-Fajfer, O. (2016). Literatura a proces rozwoju i rewitalizacja tożsamości językowej na przykładzie literatury łemkowskiej. In J. Olko, T. Wicherkiewicz, & R. Borges (Eds.), *Integral Strategies for Language Revitalization* (pp. 177-178). Warszawa, Poland: Faculty of "Artes Liberales", University of Warsaw. Retrieved from http://revitalization.al.uw.edu.pl/Content/Uploaded/Documents/integral-strategies-a91f7f0d-ae2f-4977-8615-90e4b7678fcc.pdf#page=177

DuPont, Q. (2018, May). The Cryptological Origins of Machine Translation, from al-Kindi to Weaver. (C. Mitchell, & R. Raley, Eds.) *amodern*(8), 1-20. Retrieved May 22, 2021, from http://amodern.net/article/cryptological-origins-machine-translation/

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2021). *How many languages are there in the world?* (D. M. Eberhard, G. F. Simons, & C. D. Fennig, Eds.) Retrieved June 13, 2021, from Ethnologue: Languages of the World: https://www.ethnologue.com/guides/how-many-languages

Fortson IV, B. W. (2004). *Indo-European Language and Culture.* Malden, MA, USA: Blackwell Publishing.

Google. (2021, June 8). *Language Support | Cloud Translation.* Retrieved June 13, 2021, from Google Cloud: https://cloud.google.com/translate/docs/languages

Hajlaoui, N., Kolovratnik, D., Vaeyrynen, J., Steinberger, R., & Varga, D. (2014). DCEP -Digital Corpus of the European Parliament. *Language Resources and Evaluation Conference (LREC 2014)*, (pp. 3164-3171). Reykjavik, Iceland. Retrieved June 19, 2021, from http://www.lrec-conf.org/proceedings/lrec2014/pdf/943_Paper.pdf

Horoszczak, J. (2004). *Słownik łemkowsko-polski, polsko-łemkowski.* Warszawa, Poland: Fundacja Wspierania Mniejszości Łemkowskiej Rutenika.

Jassem, W. (2003, June). Polish. *Journal of the International Phonetic Association, 33*(1), 103-107. doi:10.1017/S0025100303001191

Jónsson, H. P., Símonarson, H. B., Snæbjarnarson, V., Steingrímsson, S., & Loftsson, H. (2020). Experimenting with Different Machine Translation Models in Medium-Resource Settings. In P. Sojka, I. Kopeček, K. Pala, & A. Horák (Ed.), *Text, Speech, and Dialogue. TSD 2020. Lecture Notes in Computer Science. 12284*, p. 2. Springer, Cham. doi:10.1007/978-3-030-58323-1_10

Kerča, I. (2007). *Slovnyk Rusyn'sko-Ruskŷj* (Vol. 1). Uzhhorod, Ukraine: PolyPrynt.

Kocmi, T. (2020). CUNI Submission for the Inuktitut Language in WMT News 2020. *Proceedings of the 5th Conference on Machine Translation (WMT)*, (pp. 171–174). Association for Computational Linguistics. Retrieved June 19, 2021, from https://www.aclweb.org/anthology/2020.wmt-1.14

Kocmi, T., & Bojar, O. (2019). CUNI Submission for Low-Resource Languages in WMT News 2019. *Proceedings of the Fourth Conference on Machine Translation (WMT). Volume 2: Shared Task Papers (Day 1)*, pp.

234–240. Florence, Italy: Association for Computational Linguistics. Retrieved June 13, 2021, from https://www.aclweb.org/anthology/W19-5322.pdf

Lewis-Kraus, G. (2016, December 14). The Great A.I. Awakening (Going Neural). *The New York Times*, p. 40. Retrieved from https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

Macken, L., Prou, D., & Tezcan, A. (2020, April 23). Quantifying the Effect of Machine Translation in a High-Quality Human Translation Production Process. *Informatics, 7*(2). doi:10.3390/informatics7020012

Maximova, S., Noyanzina, O., Omelchenko, D., & Maximova, M. (2018). The Russian-speakers in the CIS countries: migration activity and preservation of the Russian language. In P. Vladimirovich (Ed.), *2018 International Scientific Conference "Investment, Construction, Real Estate: New Technologies and Special-Purpose Development Priorities" (ICRE 2018)* , *212.* Irkutsk, Russia. doi:10.1051/matecconf/201821210005

Microsoft. (n.d.). *Turn airplane mode on or off.* Retrieved June 9, 2021, from Microsoft: https://support.microsoft.com/en-us/windows/turn-airplane-mode-on-or-off-f2c2e0a1-706f-ff26-c4b2-4a37f9796df1

NATO Review. (n.d.). *About us.* Retrieved June 9, 2021, from North Atlantic Treaty Organization: https://www.nato.int/docu/review/about.html

Ng, N., Yee, K., Baevski, A., Ott, M., Auli, M., & Edunov, S. (2019, August). Facebook FAIR's WMT19 News Translation Task Submission. *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, 314-319. Florence, Italy: Association for Computational Linguistics. doi:10.18653/v1/W19-5333

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., . . . Auli, M. (2019). fairseq: A Fast, Extensible Toolkit for Sequence Modeling. *Proceedings of NAACL-HLT 2019: Demonstrations. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 48-53. Minneapolis, MN: Association for Computational Linguistics. doi:10.18653/v1/N19-4009

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 on Association for Computational Linguistics* (pp. 311-318). Philadelphia, pa: Annual Meeting of the Association for Computational Linguistics.

Post, M. (2018, September 12). A Call for Clarity in Reporting BLEU Scores. *Amazon Research*.

Rabus, A., & Scherrer, Y. (2017). Lexicon Induction for Spoken Rusyn – Challenges and Results. *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, (pp. 27-32). Valencia, Spain.

Scherrer, Y., & Rabus, A. (2017). Multi-source morphosyntactic tagging for Spoken Rusyn. *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects* (pp. 84-92). Valencia, Spain: Association for Computational Linguistics. doi:http://dx.doi.org/10.18653/v1/W17-1210

Scherrer, Y., & Rabus, A. (2019, September). Neural morphosyntactic tagging for Rusyn. (R. Mitkov, Ed.) *Natural Language Engineering, 25*(5), pp. 633-650. doi:10.1017/S1351324919000287

Shea, J. (2016, December 5). What can we learn today from the 'three wise men'? *NATO Review*. Retrieved May 26, 2021, from https://www.nato.int/docu/review/articles/2016/12/05/what-can-we-learn-today-from-the-three-wise-men/index.html

UNESCO Ad Hoc Expert Group on Endangered Languages. (2003). Language Vitality and Endangerment. *International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages.* Paris: UNESCO. Retrieved June 19, 2021, from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/Language_vitality_and_endangerment_EN.pdf

Vasmer, M. J. (n.d.). *Etimologičeskyj Slovar' Russkogo Jazyka.* (O. N. Trubačëv, Trans.) Moscow: AST (Original work published 1950).

Watral, M. (2015, February). Rewitalizacja Łemków. *Znak*(717), 38-44. Retrieved August 24, 2021, from https://www.miesiecznik.znak.com.pl/7172015marta-wartalrewitalizacja-lemkow/

Watral, M. (2016). Postawy względem języka łemkowskiego – wzór i jego realizacja. In J. Olko, T. Wicherkiewicz, & R. Borges (Eds.), *Integral Strategies for Language Revitalization* (pp. 221-260). Warsaw, Poland: Faculty of "Artes Liberales", University of Warsaw. Retrieved August 24, 2021, from http://revitalization.al.uw.edu.pl/Content/Uploaded/Documents/integral-strategies-a91f7f0d-ae2f-4977-8615-90e4b7678fcc.pdf#page=243

Ziemski, M., Junczys-Dowmunt, M., & Pouliquen, B. (2016). The United Nations Parallel Corpus v1.0. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, (pp. 3530–3534). Portorož, Slovenia. Retrieved from https://www.aclweb.org/anthology/L16-1561