# Say It Right: AI Neural Machine Translation Empowers New Speakers To Revitalize Lemko

Petro Orynycz<sup>[0000-0003-3094-9156]</sup>

LemkoTran.com, Edgewater, Maryland 21037, USA p@orynycz.com

**Abstract.** Artificial-intelligence powered neural machine translation might soon resuscitate endangered languages by empowering new speakers to communicate in real time using sentences quantifiably closer to the literary norm than those of native speakers, and starting from day one of their language reclamation journey. While Silicon Valley has been investing enormous resources into neural translation technology capable of superhuman speed and accuracy for the world's most widely used languages, 98% have been left behind, for want of corpora: neural machine translation models train on millions of words of bilingual text, which simply do not exist for most languages, and cost upwards of a hundred thousand United States dollars per tongue to assemble.

For low-resource languages, there is a more resourceful approach, if not a more effective one: transfer learning, which enables lower-resource languages to benefit from achievements among higher-resource ones. In this experiment, Google's English-Polish neural translation service was coupled with my classical, rule-based engine to translate from English into the endangered, low-resource, East Slavic language of Lemko. The system achieved a bilingual evaluation understudy (BLEU) quality score of 6.28, several times better than Google Translate's English to Standard Ukrainian (BLEU 2.17), Russian (BLEU 1.10), and Polish (BLEU 1.70) services. Finally, the fruit of this experiment, the world's first English to Lemko translation service, was made available at the web address www.LemkoTran.com to empower new speakers to revitalize their language.

New speakers are key to language revitalization, and the power to "say it right" in Lemko is now at their fingertips.

Keywords: Human-Centered AI, Language Revitalization, Lemko.

Please cite as: Orynycz, P. (2022). 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. Springer, Cham. <u>https://doi.org/10.1007/978-3-031-05643-7\_37</u>

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 <a href="https://doi.org/10.1007/978-3-031-05643-7\_37">https://doi.org/10.1007/978-3-031-05643-7\_37</a>. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use: <a href="https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms">https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms</a>.

# 1 Introduction

### 1.1 Problems

This experiment aims to contribute at the local level to the global challenge of language loss, which may be occurring at the rate of one per day, with as few as one tongue in ten set to survive [1, p. 1329]. At press time, SIL International's *Ethnologue* uses Lewis and Simons' 2010 Expanded Graded Intergenerational Disruption Scale to estimate that 3,018 languages are endangered [2], which is 43% of the 7,001 individual living ones tallied at press time in International Organization for Standardization standard ISO 639-3 [3]. Meanwhile, Google Translate only serves 108 [4], and Facebook, 112 [5], which is a start. Nevertheless, one less language is now underserved, as the fruit of this experiment has been deployed to a web server as a public translation service.

New, artificial intelligence technologies beckon with the promise of an aid that instantly compensates for language loss via human-computer interaction. In my previous experiment, next-generation neural engines achieved higher quality scores translating from Russian and Polish into English than the human control [6, p. 9]. Meanwhile, Facebook and Google<sup>1</sup> have invested enormous resources into delivering better-thanhuman automatic translation systems at zero cost to consumer.

Superhuman artificial intelligence does not come cheap: training neural language models requires bilingual corpora with wordcounts in the hundreds of thousands, and ideally, millions, which would cost hundreds of thousands of dollars to translate, sums beyond the means of most low-resource language communities. Fortunately, this experiment shows that there are more resourceful and effective ways to respond to the challenge of creating translation aids for revitalizing endangered languages in low-resource settings.

#### 1.2 Work So Far

I built the world's first Lemko to English machine translation system and have made it available to the public. Its objective translation quality scores have been improving: the engine achieved a bilingual evaluation understudy (BLEU) score of 14.57 in the summer of 2021, as presented to professionals at the National Defense Industrial Association's Interservice/Industry Training, Simulation and Education Conference and published in its proceedings [6]. For reference, I scored BLEU 28.66 as a human translator working in field conditions, cut off from the outside world. By the autumn of 2021, the engine had reached BLEU 15.74, as reported to linguists, academics, and the wider community at an unveiling event hosted by the University of Pittsburgh.<sup>2</sup>

<sup>&</sup>lt;sup>1</sup> Disclosure: I work as a paid Russian, Polish, and Ukrainian linguist and translation quality control specialist for the Google Translate project; headquarters are in San Francisco.

<sup>&</sup>lt;sup>2</sup> Disclosure: the event was sponsored by the Carpatho-Rusyn Society (Pennsylvania), and I was paid by the University of Pittsburgh for my presentation.

### 1.3 System Under Study

Lemko is a definitively to severely endangered [6, p. 3, 7, pp. 177-178], low-resource [8], officially recognized minority language [9] presumably indigenous to transborder highlands south of the Cracow, Tarnów, and Rzeszów metropolitan areas; historical demarcating isoglosses will hopefully be the topic of a future paper. Poland's census bureau tallied 6,279 residents for whom Lemko was a language "usually used at home" (even if in addition to Polish) in 2011 [10, p. 3], a 12% increase from the 5,605 for whom Lemko was a "language spoken most often at home" in 2002 [11, p. 6, 12, p. 7]. At press time, the results of a fresh count are being tabulated.

Lemko is classifiable as an East Slavic language as it fits the customary genetic structural feature criteria, the most significant of which is pleophony [13, p. 20], whereby a vowel is assumed to have arisen in proto-Slavic sequences of consonant *C* followed by mid or low vowel V (\*e, or \*o, with which \*a had merged [14, p. 366]), followed by liquid R (that is, \*l or \*r), followed by another consonant *C*, that is, CVRC > CVRVC. To illustrate, compare the Old English word for "melt", *meltan* (CVRC) [15, p. 718] to its putative Lemko cognate *molódyj* [16, p. 92, 17, p. 150] (CVRVC), meaning "young". Other East Slavic cognates include Ukrainian *molodýj* and Russian *molodój* [17], both exhibiting a vowel after the liquid (CVRVC). Meanwhile, West Slavic languages lack a vowel before the liquid; compare Polish *mlody* and Slovak *mladý* (both CRVC) [17]. Further afield, kinship has been posited for other words translatable as "mild", including Sanskrit *mrdú* (CRC) [18, p. 830] and Latin *mollis* (CVRC if from \**moldvis*) [15, 17, 19, p. 323].

How well Lemko meets customary, modern Ukrainian genetic structural feature criteria was not evaluated in this experiment. However, similarity between Lemko and Standard Ukrainian was quantified, for the first time in print of which I am aware. Below, my Lemko engine scored BLEU 6.28, nearly three times the score of Google Translate's Ukrainian at BLEU 2.17. Further experiments could be performed for the purposes of quantification of similarity between Lemko, Standard Ukrainian, Polish, and Rusyn as codified in Slovakia, as well as a fresh take on the typological classification of Lemko.

The quantity and quality of resources have been improving, as has resourcefulness empowered by technology. All known bilingual corpora, comprising fewer than seventy thousand Lemko words, were mustered for this experiment. I have been cleaning a bilingual corpus of transcriptions of interviews conducted with native speakers in Poland and my translations into English, which a United States client paid me to perform and permitted me to use. I am also compiling monolingual corpora, which total 534,512 words at press time.

### 1.4 Hypothesis

Based on my subjective impression as a professional translator that Lemko native speakers interviewed in Poland were more likely to use words with obvious Polish cognates than Standard Ukrainian ones, I hypothesized that, all else being equal, a machine could be configured to translate into Lemko from English and achieve BLEU objective quality scores higher than those of Google Translate's Ukrainian and Russian services.

### 1.5 Predictions

**Lemko Translation System.** I predicted that the aforementioned translation system would achieve a BLEU score of 15 translating into Lemko from English against the bilingual corpus.

### **Google Translate.**

*English to Ukrainian service.* I predicted that Google Translate's English to Ukrainian service would achieve a BLEU score of 10 against the bilingual corpus.

*English to Russian service*. I predicted that Google Translate's English to Russian service would achieve a BLEU score of 1 against the bilingual corpus.

### 1.6 Methods and Justification

In the interest of speed, resource conversation, and ruggedizability, a laptop computer discarded as obsolete by my employer was configured to translate into Lemko and make calls to the Google Cloud Platform Google Translate service, as well as configured to evaluate said translations using the industry standard BLEU metric.

### 1.7 Principal Results

The English to Lemko translation system achieved a cumulative BLEU score of 6.28431824990417. Meanwhile, Google Translate's Ukrainian service scored BLEU 2.16830846776652, its Russian service BLEU 1.10424105952048, and the control of Polish transliterated into the Cyrillic alphabet BLEU 1.70036447680114.

# 2 Materials and Methods

The above hypothesis was tested by calculating BLEU quality scores for each translation system set up in the manner detailed below.

### 2.1 Setup

**Hardware.** The experiment was conducted on an HP Elitebook 850 G2 laptop with a Core i7-5600U 2.6GHz processor, and 16 gigabytes of random-access memory. It had been discarded by my employer as obsolete and listed for sale at USD 450 at time of press.

*Configuration.* In the basic input/output system (BIOS) menu, the device was configured to enable Virtualization Technology (VTx).

**Operating System.** Windows 10 Professional 64 bit had been installed on bare metal. It was ensured that Virtual Machine Platform and Windows Subsystem for Linux Windows features were enabled. Next, the WSL2 Linux kernel update for x64 machines (wsl\_update\_x64.msi) available from Microsoft at https://aka.ms/wsl2kernel was installed.

**Software.** The Docker Desktop for Windows version 4.4.3 (73365) installer was downloaded from https://www.docker.com/get-started and run with the option to Install required Windows components for WSL 2 selected.

**Packages.** The experiment depended on the below packages from the Python Package Index.

*SacreBLEU*. Version 2.0.0 was installed using the Python package documented at the following universal resource locator (URL):

https://pypi.org/project/sacrebleu/2.0.0/

Google Cloud Translation API client library. Version 2.0.1 was installed using the Python package documented at the universal resource locator (URL) https://pypi.org/project/google-cloud-translate/2.0.1/

The above dependencies were specified in the requirements file as follows:

google-cloud-translate==2.0.1
sacrebleu==2.0.0

### Container.

*Build.* The experiment was run in a Docker container featuring the latest version of the Python programming language, which was version 3.10.2 at the time, running on the Debian Bullseye 11 Linux operating system of AMD64 architecture, of Secure Hash Algorithm 2 shortened digest bcb158d5ddb6, obtainable via the following command:

```
docker pull
python@sha256:bcb158d5ddb636fa3aa567c987e7fcf61113307820d
466813527ca90d60fedc7
```

*Runtime.* The container was configured to save raw experiment data files to a local bind mounted volume.

### **Translation Quality Scoring.**

Translation quality scores were calculated according to the BLEU metric using version 2.0.0 of the *SacreBLEU* tool invented by Post [20].

Case sensitivity. The evaluation was performed in a case-sensitive manner.

**Tokenization.** Segments were tokenized using version 13a of the Workshop on Statistical Machine Translation standard scoring script metric internal tokenization procedure.

**Smoothing Method.** The smoothing technique developed at the National Institute of Standards and Technology by United States Federal Government employees for their Multimodal Information Group BLEU toolkit, being the third technique described by Chen and Cherry [21, p. 363], was employed by default.

Signature. The above settings produced the following signature:

```
nrefs:1|case:mixed|eff:no|tok:13a|smooth:exp|ver-
sion:2.0.0
```

Calibration. Configured as above, the machine produces the following output:

English source	Everything was	there.
Lemko reference and transliteration	Вшытко там было.	Všŷtko tam bŷlo.
Lemkotran.com hypothesis and transliteration	Вшытко там было.	Všŷtko tam bŷlo.
Score	BLEU = 100.00 100.0/100.0/100.0/100.0 (BP = 1.000 ratio = 1.000 hyp_len = 4 ref_len = 4)	

Segment 1031.

*Explanation*. The hypothesis segment was identical to the reference one and the machine achieved a perfect score of BLEU 100.

Segment 17	9.

English source	I don't remember what year.	
Lemko reference and transliteration	Не памятам в котрым році.	Ne pamjatam v kotrŷm roci.
Lemkotran.com hypothesis and transliteration	Ні памятам, в котрым році.	
Score	BLEU = 43.47 71.4/50.0/40.0/25.0 (BP = 1.000 ratio = 1.167 hyp_len = 7 ref_len = 6)	

*Explanation.* The hypothesis was different from the reference by two characters. The machine mistranslated the particle negating the verb, using the word for "no" (*ni*) instead of the expected word for "not" (*ne*). This has since been largely fixed. The machine also added a comma after *pamjatam*, which means "I remember". That dropped the score from what would have been a perfect score of 100 to 43.47.

**Control.** As the corpus is based on interviews conducted in Poland, translations into Polish were used as a control. They were transliterated into the Cyrillic alphabet by reversing the rules for transliterating Lemko names established by Poland's Ministry of the Interior and Administration [22, p. 6564]. Polish nasal vowels were decomposed into a vowel plus a nasal stop, except before approximants, where they were directly denasalized. Word finally, the front nasal vowel /ę/ was simply denasalized, and the back one /ą/ was transliterated as if followed by a dental stop.

# 3 Results

The engine available to the public at www.LemkoTran.com took first place with a cumulative translation quality score of BLEU 6.28, nearly three times that of the runner-up, Google Translate's English-Ukrainian service (BLEU 2.17). Next was its English-Polish service (BLEU 1.70), with its English-Russian service in last place (BLEU 1.10).



Table 1. English to Lemko Translation Quality: LemkoTran.com versus Google Translate

## 3.1 Results by machine translation service

**Control.** When transliterated into the Cyrillic alphabet, Google Translate's translations into Standard Polish achieved a corpus-level BLEU score of 1.70. Samples of its performances are as follows:

segment 2174.		
English source	We had still been in Izby, right.	
Lemko reference and transliteration	То мы іщы были в Ізбах, так.	To mŷ iščŷ bŷly v Izbach, tak.
Polish hypothesis and transliteration	Билісьми єще в Ізбах, так.	Byliśmy jeszcze w Izbach, tak.
Score	BLEU = 46.20	

Segment 2174.

|--|

English source	And that's what it	's all about.
Lemko reference and transliteration	I о то ходит.	I o to chodyt.
Polish hypothesis and transliteration	I о то власьнє ходзі.	I o to właśnie chodzi.
Score	BLEU = 32.47	

C	<b>n</b> 1	7	
Noomont	11	/	
oczmeni		<i>'</i> •	

English source	And that's what it's all about.	
Lemko reference and transliteration	Так мі повіл.	Tak mi povil.
Polish hypothesis and transliteration	Так мі повєдзял.	Tak mi powiedział.
Score	BLEU = 35.36	

**Hybrid English-Lemko Engine.** The engine freely available to the public at the URL www.LemkoTran.com achieved a corpus-level BLEU score of 6.28.

Segment	1031

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~		
English source	Everything was there.	
Lemko reference and transliteration	Вшытко там было.	Všýtko tam býlo.
Lemkotran.com hypothe- sis and transliteration	Вшытко там было.	Všýtko tam býlo.
Score	BLEU = 100.00	

Segment	1445.

English source	But that officer took that medal and said,	
Lemko reference and transliteration	Але тот офіцер взял тот медаль і повідат:	Ale tot oficer vzial tot medal' i povidat:
Lemkotran.com hypothe- sis and transliteration	Але тот офіцер взял тот медаль і повіл:	Ale tot oficer vzial tot medal' i povil:
Score	BLEU = 75.06	

# Segment 217.

English source	That's what he said to me.		
Lemko reference and transliteration	Так мі повіл.	Tak mi povil.	
Lemkotran.com hypothe- sis and transliteration	Так мі повіл.	Tak mi povil.	
Score	BLEU = 100.00		

**Ukrainian.** Google Translate's translations into Standard Ukrainian achieved a corpuslevel BLEU score of 2.35.

Segment 2117.		
English source	Where and when?	
Lemko reference and transliteration	Де і коли?	De i koly?
Ukrainian hypothesis and transliteration	Де і коли?	De i koly?
Score	BLEU = 100.00	

Segment 2419.

Segment	1096.

English source	We were there for three months.		
Lemko reference and transliteration	Там зме были три misiaci.		
Ukrainian hypothesis and transliteration	Ми були там три місяці.	My buly tam try misjaci.	
Score	BLEU = 30.21		

	Segment	2513.
--	---------	-------

English source	Well, here to the west.	
Lemko reference and transliteration	Но то ту на захід.	No to tu na zachid.
Ukrainian hypothesis and transliteration	Ну, тут на захід.	Nu, tut na zachid.
Score	BLEU = 30.21	

**Russian.** Google Translate's English to Russian service achieved a corpus-level BLEU score of 1.10.

Segment	432.

English source	Nobody knew.	
Lemko reference and transliteration	Нихто не знал.	Nychto ne znal.
Russian hypothesis and transliteration	Никто не знал.	Nikto ne znal.
Score	BLEU = 59.46	

Segment	2751
Segment	2/01.

English source	What did they expel us for?	
Lemko reference and transliteration	За што нас выгнали?	Za što nas vŷhnaly?
Russian hypothesis and transliteration	За что нас выгнали?	Za čto nas vyg- nali?
Score	BLEU = 42.73	

S	egi	mei	nt	21	64	l <u>.</u>	

English source	Brother went off to war.	
Lemko reference and transliteration	Брат пішол на войну.	Brat pišol na vojnu.
Russian hypothesis and transliteration	Брат ушел на войну.	Brat ušel na vojnu.
Score	BLEU = 42.73	

## 4 Discussion

The Lemko translation system corpus-level BLEU score of 6.28 indicates that while there is much still to be done, things are on track. The Standard Russian score of BLEU 1.10 indicates that Lemko is less similar to Russian than Polish (BLEU 1.70). Perhaps using pre-revolutionary orthography could boost Russian's score, but that would be an expensive experiment with little obvious benefit.

The transliterated Standard Polish control similarity score of BLEU 1.70 indicates less interference from the dominant language in Poland than might be expected. It would be interesting to redesign the experiment where a handful of computationally inexpensive and obvious sound correspondences (for example, denasalization of \*e to /ja/ and \*o to /u/, retraction of \*i to /y/, and change of \*g to /h/ [23]) were applied to Polish to see if it then scored higher than Standard Ukrainian.

In summary, Lemko has been synthesized in the lab and the power to produce it placed in the hands of speakers both new and native. After a thorough engine overhaul and glossary ramp-up, the next step is to objectively measure, and if feasible, have speakers subjectively rate, the quality of synthetic Lemko versus that produced by native speakers. The day when new speakers of low-resource languages can use machine translation to start communicating in their language overnight is closer, as is the day the Lemko language joins the ranks of those previously endangered, but now revitalized. Acknowledgements. I would like to thank my colleague Ming Qian of Peraton Labs for inspiring me to conduct this experiment, and Brian Stensrud of Soar Technology, Inc. for introducing us, as well as his encouragement.

I would also like to thank my friend Corinna Caudill for her encouragement and personal interest in the project, as well as for introducing me to Carpatho-Rusyn Society President Maryann Sivak of the University of Pittsburgh, whom I would like to thank for the opportunity to present my work.

I would also like to thank Maria Silvestri of the John and Helen Timo Foundation for conducting interviews with Lemko native speakers and donating the transcripts and my translations of them to research and development.

I would like to Achim Rabus of the University of Freiburg and Yves Scherrer of the University of Helsinki for their interest in the project and ideas.

I would also like to thank Myhal' Lŷžečko of the minority-language technology blog InterFyisa for his early interest in the project and community outreach.

I would also like to thank fellow son of Zahoczewie Marko Łyszyk for his interest in the project and community outreach.

Finally, I would like to thank my co-author and Antech Systems Inc. colleague Tom Dobry for his encouragement and guidance.

#### References

- Graddol, D.: The future of language. Science, 303(5662), 1329-1331 (2004). https://doi.org/10.1126/science.1096546
- Eberhard, D. M., Simons, G. F., & Fennig, C. D.: Ethnologue: Languages of the World, SIL International. Twenty-fourth edition. SIL International, Dallas (2021). Online version: How many languages are endangered?, https://www.ethnologue.com/guides/how-many-languages-endangered, last accessed 2022/02/11.
- 3. ISO 639 Code Tables, https://iso639-3.sil.org/code\_tables/639/data, last accessed 2022/02/11.
- 4. Language support, https://cloud.google.com/translate/docs/languages, last accessed 2022/02/11.
- 5. Select language, https://m.facebook.com/language.php, last accessed 2022/02/11.
- 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 (I/ITSEC) 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
- Duć-Fajfer, O.: Literatura a proces rozwoju i rewitalizacja tożsamości językowej na przykładzie literatury lemkowskiej. In: Olko, J., Wicherkiewicz, T., Borges, R. (eds.), Integral Strategies for Language Revitalization, pp. 175–200. First edition. Faculty of "Artes Liberales", University of Warsaw, Warsaw (2016).
- Scherrer, Y., Rabus, A.: Neural morphosyntactic tagging for Rusyn. In: Mitkov, R., Tait, J., Boguraev, B. (eds.), Natural Language Engineering, 25(5), 633–650. Cambridge University Press, Cambridge (2019). https://doi.org/10.1017/S1351324919000287

- Reservations and Declarations for Treaty No.148 European Charter for Regional or Minority Languages (ETS No. 148), https://www.coe.int/en/web/conventions/full-list?module=declarations-by-treaty&numSte=148&codeNature=1&codePays=POL, last accessed 2022/02/11.
- Formularz indywidualny, https://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultstronaopisowa/5781/1/1/nsp\_2011\_badanie\_\_pelne\_wykaz\_pytan.pdf, last accessed 2022/02/11.
- Narodowy Spis Powszechny Ludności i Mieszkań 2002 r. z 20 maja (formularz A) https://stat.gov.pl/gfx/portalinformacyjny/userfiles/\_public/spisy\_powszechne/nsp2002form-a.pdf, last accessed 2022/02/11.
- IV Raport dotyczący sytuacji mniejszości narodowych i etnicznych oraz języka regionalnego w Rzeczypospolitej Polskiej – 2013, http://mniejszosci.narodowe.mswia.gov.pl/download/86/14637/TekstIVRaportu.pdf, last accessed 2022/02/11.
- Vaňko, J.: The Language of Slovakia's Rusyns. East European Monographs, New York (2000).
- 14. Forston, B., IV: Indo-European Language and Culture. Blackwell Publishing, Oxford (2004).
- 15. Pokorny, J.: Indogermanisches etymologisches Wörterbuch, Bern, 1959.
- 16. Horoszczak, J.: Słownik lemkowsko-polski, polsko-lemkowski. Rutenika, Warsaw (2004).
- 17. Vasmer, M. Russisches etymologisches Wörterbuch. Zweiter Band. Carl Winter, Universitätsverlag, Heidelberg (1955).
- Monier-Williams, M.: A Sanskrit-English Dictionary Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages, The Clarendon Press, Oxford (1899).
- Derksen, R.: Etymological Dictionary of the Slavic Inherited Lexicon. In: Lubotsky, A. (ed.) Leiden Indo-European Etymological Dictionary Series, vol. 4, Koninklijke Brill, Leiden (2008).
- 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://aclanthology.org/W18-63
- Chen B., Cherry, C.: A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU. In: Proceedings of the Ninth Workshop on Statistical Machine Translation, pp. 362–367. Association for Computational Linguistics, Baltimore (2014). http://dx.doi.org/10.3115/v1/W14-33
- 22. Ministerstwo Spraw Wewnętrznych i Administracji: Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 30 maja 2005 r. w sprawie sposobu transliteracji imion i nazwisk osób należących do mniejszości narodowych i etnicznych zapisanych w alfabecie innym niż alfabet łaciński. In: Dziennik Ustaw Nr 102, pp. 6560–6573. Rządowe Centrum Legislacji, Warsaw (2005).
- Shevelov, G.: On the Chronology of H and the New G in Ukrainian. In: Harvard Ukrainian Studies, vol. 1, no. 2, pp. 137–152. Harvard Ukrainian Research Institute, Cambridge (1977). https://www.jstor.org/stable/40999942