

AD: KRITICKÝ NÁHĽAD NA AUTOMATICKÝ PREKLADAČ DO ŠTÚROVSKEJ SLOVENČINY¹

Kristína Bobeková – Jana Levická – Michaela Majerčíková

BOBEKOVÁ, K. – LEVICKÁ, J. – MAJERČÍKOVÁ, M. (2024): Ad: Critical View on the Automatic Translator into Štúr's Slovak. In: *Slovenská reč*, 89/1, 164 – 171.

Abstract: The article offers constructive criticism not only of the specialized tool “ludevít“, machine translation into Štúr's Slovak, but also the recently published critical overview of S. Habijanec (2023) of the aforementioned tool. The authors aim to refute some distorted claims that were published in *Slovenská reč* – the intention of creating “ludevít“ and its limits, misleading claims about the organization and purpose of the literary competition *Báseň v štúrovčine*, and finally they discuss the (un)ethical usage of pupils' and students' works for scientific analysis. For these reasons, it can be assumed that not only the absence of more thorough preparation by S. Habijanec, but also the absence of addressing the author of the tool R. Garabík – for the sake of possible cooperation in order to improve the tool – prevented the critical view on “ludevít“ tool, published in *Slovenská reč* in 2023, from fulfilling its potential.

Keywords: Ľudovít Štúr, Štúr's Slovak, Slovak language, machine translation, critical view on the automatic translator.

Článok v skratke:

- Príspevok reaguje na nekorektné, skresľujúce a zavádzajúce informácie uverejnené v článku S. Habijanca na stránkach časopisu *Slovenská reč* v roku 2023.
- V príspevku je ozrejmený autorský zámer vytvorenia nástroja „ludevít“ spolu s kritickým náhľadom na jeho limity.
- Príspevok uvádza na pravú mieru informácie o organizácii a celi dvoch ročníkov literárnej súťaže *Báseň v štúrovčine*.
- Autorky príspevku polemizujú nad (ne)etickosťou voľby materiálu na vedeckú analýzu v podobe básní najmä žiakov a študentov rôznych vekových kategórií.

1. ÚVOD

Týmto príspevkom chceme reagovať na článok S. Habijanca *Kritický náhľad na automatický prekladač do štúrovskej slovenčiny*, ktorý bol zverejnený v 1. čísle časopisu *Slovenská reč* v roku 2023. Cieľom tohto textu nie je vyjadriť sa ku všet-

¹ Príspevok vznikol ako reakcia na článok *Kritický náhľad na automatický prekladač do štúrovskej slovenčiny* (Habijanec 2023) v rámci projektu *Tvorba a rozvoj Slovenského národného korpusu (V. etapa)*, financovaného Ministerstvom školstva, vedy, výskumu a športu SR, Ministerstvom kultúry SR a Slovenskou akadémiou vied.

kým bodom kritiky využívania nástroja „ludevít“. Naším zámerom je uviesť na pravú mieru niektoré skresľujúce tvrdenia: 1) autorský zámer vytvorenia nástroja „ludevít“ a jeho limity; 2) organizácia literárnej súťaže *Báseň v štúrovčine*; 3) cieľ literárnej súťaže *Báseň v štúrovčine* a 4) básne žiakov a študentov ako materiál na vedeckú analýzu.

2. AUTORSKÝ ZÁMER VYTVORENIA NÁSTROJA „LUDEVÍT“ A JEHO LIMITY

Nástroj „ludevít“² je v prvom rade ukážkou možností automatického prepisu textu z jazyka A do jazyka B, v tomto prípade zo súčasnej slovenčiny do jej historického štúrovského variantu, na základe slovníka a sady prekladových pravidiel, tzn. strojový preklad založený na pravidlách. Treba zároveň zdôrazniť, že zvolený typ mal dosiahnuť čo najvyššiu úspešnosť prekladu bez väčšieho rozsahu ručnej práce. Ašpirovať na presné a spoľahlivé výstupy pri jeho využívaní nemožno v prvom rade s ohľadom na rok jeho zverejnenia, pretože nástroj vznikol v čase, keď strojový preklad vo všeobecnosti slúžil skôr len na základné pochopenie textu. Nástroj sa od svojho prvého zverejnenia vyvíjal len minimálne^{3[1], [2]} a slúžil ako ukážka technologických možností jednoduchého pravidlového prevodu textu.

Autorským zámerom bolo vytvoriť **prototyp**⁴ softvérového prostredia, webového rozhrania a knižníc pre jednoduchú transformáciu textu na základe ortografických a lexikálnych pravidiel. Toto prostredie bolo koncipované ako univerzálne, nezávislé od jazyka/jazykov, ktorých transformáciu by poskytovalo. Vytvorenie prekladača do štúrovskej slovenčiny bolo prvou aplikáciou fungujúcou v tomto prostredí. Neskôr pribudli ďalšie aplikácie, ktoré využívajú navrhnuté postupy – experimentálne webové rozhranie poskytujúce prístup k štatistickému strojovému prekladu nástroja MOSES v rámci projektu EuroMatriPlus^{3]}, a takisto pomerne populárnemu rozhraniu poskytujúcemu transliteráciu slovenčiny, češtiny alebo chorvátčiny do hlaholiky^{4]}. Nástroj „ludevít“ mal poskytnúť skôr vytvorenie predstavy o štúrovskej slovenčine, nie stricto sensu Štúrovej, a hlavne vtipným⁵ spôsobom, atraktívnejším pre širšiu verejnosť, propagovať aktivity Jazykovedného ústavu Ľ. Štúra SAV. Podobné aktivity zahŕňali digitalizáciu diel *Nauka reči Slovenskej a Nárečja Slovensku alebo Potreba písania v tomto nárečí*.

Štúrovský variant slovenčiny sa od súčasnej spisovnej slovenčiny na prvý pohľad líši ortografiou – napríklad absenciou grafém *y*, *ô*, *l'*, a *í*, inou realizáciou dvojhlasok, odlišným značením mäkkosti spoluhlások. Nástroj „ludevít“ zachytáva tieto rozdielnosti a konvertuje ich deterministicky, tzn., že grafému *y* nahrádza

² Nástroj „ludevít“ bol vytvorený R. Garabíkom a sprístupnený verejnosti v roku 2006.

³ Zmeny boli minimálne, väčšinou len technické alebo lexikálne.

⁴ Na zdôraznenie kľúčových informácií používame v texte **tučné písmo**.

⁵ Necháme na posúdenie používateľom, do akej miery sa tento zámer podarilo naplniť.

grafémou *i* (*účely* – *účeli*), napríklad dvojhlásky *ia* a *ie* nahrádza *ja*, *je* (*dialka* – *djalka*, *rieka* – *rjeka*), k mäkkým spoluhláskam *d*, *t*, *n* povinne pričleňuje mäkkosť v podobe mäčkčov (*deti* – *d'eti*, *ticho* – *t'icho*, *nikto* – *ňikdo*). Z týchto dôvodov nikdy nebolo možné nástroj „ludevít“ považovať za spoľahlivú „prekladateľskú“ pomôcku a očakávať od neho gramaticky správny výsledok, pretože jeho cieľom nie je ponúkať korektný prepis do štúrovskej slovenčiny. Pôvodný cieľ nástroja ostáva zachovaný dodnes, a to, že má slúžiť najmä na pobavenie a popularizáciu vedeckého výskumu jazyka. V tomto duchu prichádza aj spätná reakcia od používateľov daného nástroja.

V prvom rade si však kladieme otázku, prečo S. Habijanec nesiahol pri svojej kritike po publikovanom článku R. Garabíka (2009) a kompletne dostupných prekladových tabuľkách nástroja „ludevít“^[5]. Zdrojový kód, vrátane spomínaných tabuliek, bol zverejnený od roku 2006. Ak by autor kritiky nahliadol do spomenutého článku, mal by k dispozícii komplexný opis technickej realizácie nástroja „ludevít“, spolu s popisom funkcií programu a relatívne podrobným – s uspokojivým množstvom príkladov – a najmä kritickým náhľadom na limity nástroja. R. Garabík (2009) uvádza vo svojom článku komplexný prehľad nedostatkov nástroja „ludevít“, ktoré možno rozdeliť na dve skupiny.

Prvá skupina zastrešuje tzv. teoreticky nepodstatné nedostatky, ktoré je možné odstrániť aplikovaním ľudského vkladu v podobe ručnej práce. Do tejto skupiny patrí najmä malý rozsah prekladového slovníka a nedostatky v slovníkoch nielen na lexikálnej, ale aj na ortografickej úrovni. Tie možno odstrániť dôkladným skontrolovaním slovníkov a malý rozsah slovnej zásoby zase jej doplnením. Nevyhnutosť ľudského vkladu je pri prvej skupine nedostatkov takmer kľúčová a v tomto smere bol nástroj v priebehu rokov vylepšovaný a upravovaný.

Druhá skupina tzv. teoreticky podstatných nedostatkov vyplýva buď priamo z návrhu prekladového systému, alebo z vlastností samotnej spisovnej slovenčiny a jej historického štúrovského variantu a vzťahu medzi nimi. Autor nástroja pri druhej skupine nedostatkov uvádza najzávažnejšie nedostatky:

1. Absencia kontextového prekladu – prekladový systém berie do úvahy iba jedno konkrétne slovo a nie širší kontext. Práve táto skutočnosť zabráňuje napríklad možnosti lexikálneho prekladu s iným gramatickým rodom, tzn., že nie je možné súčasne preložiť napríklad adjektíva a slovesá – v menosloví – tak, aby ostala zachovaná zhoda v gramatickom rode. Prekladový systém má len jedinú výnimku pri slove *Bratislava*, ktoré prepisuje ako *Prešporok* – spolu s príslušnou paradigmou podstatného mena. Autor nástroja sa rozhodol v tomto prípade pre výnimku z pragmatických dôvodov, pretože ide o veľmi známy historický variant pomenovania *Bratislavy* aj z hľadiska používania v rôznych textoch. V niektorých prípadoch sa môžu vyskytnúť chyby v nesúlade gramatického rodu – resp. chybná kongruencia – prídavného mena pri tejto

výnimke, ktoré stále predstavujú prijateľnejší nedostatok ako ponechanie tvaru „Braťislava“, ktorý nemal jednoznačné používanie v období štúrovčiny.⁶

2. Nerozlíšiteľná homonymia v spisovnej slovenčine – R. Garabík uvádza najčastejšie príklady v podobe prídavných mien neutrálneho gramatického rodu v nominatíve a akuzatíve v singulári (koncovka *-uo*) a prídavných mien ženského a neutrálneho gramatického rodu v pluráli (koncovka *-je*), ktoré v spisovnej slovenčine majú rovnakú koncovku (koncovka *-é*, prípadne koncovka *-e*, ak bola predchádzajúca slabika dlhá). Bez morfolologickej analýzy nástroj nedokáže použiť správny variant.

3. Absencia syntaktickej roviny – historický štúrovský variant slovenčiny sa od súčasnej spisovnej slovenčiny líši v istej miere aj inou skladbou viet. Prekladový systém nástroja „ludevít“ neobsahuje žiadne prostriedky ani na syntaktickú analýzu originálu a ani na úpravu „prekladu“.

Na základe vyššie uvedených zdrojov s komplexnými informáciami by nemusel S. Habijanec čeliť dohadom o tom, že „ako [...] základný problém sa javí odlišná povaha grafických sústav štúrovčiny a súčasnej slovenčiny, ktoré táto v podstate veľmi jednoduchá aplikácia nedokáže preklenúť. Pravidlá pretvárania boli očividne koncipované tak, že na základe malého počtu príkladov sa vygenerovalo pravidlo, ktoré sa automaticky uplatňuje aj pri nenáležitých slovách výlučne na základe grafickej totožnosti“ (2023, s. 121). Zároveň podotýkame, že autor kritiky sa mýli – nástroj „ludevít“ takto nevznikol, pravidlá boli vytvorené ručne na základe publikovaných pravidiel pravopisu (Štúr 1846) a nevznikali automaticky na základe malého počtu príkladov. Hoci autor kritiky trafil správne jednu z možností, vo všeobecnosti je to použiteľný prístup – takto napríklad funguje štatistický prekladač/modernizátor historickej slovinčiny do súčasného jazyka (Scherrer – Erjavec 2016), hoci tento je trénovaný na viac než „malom počte“ príkladov.

Prekladový systém nástroja „ludevít“ umožňuje **základný** prepis zo súčasnej slovenčiny do jej historického štúrovského variantu na ortografickej a čiastočne aj na lexikálnej úrovni. Systém a ani jeho ponúkané „preklady“ nie sú dokonalé a od pôvodného historického štúrovského variantu sa odlišujú najmä v syntaktickej skladbe viet a v lexike. Napriek všetkým limitom nemožno uprieť nástroju „ludevít“ jeho popularizačné plusy v podobe demonštrácie v čase vzniku aktuálnych technologických možností jednoduchého pravidlového prevodu textu a jednoduchej ukážky štúrovčiny.

3. ORGANIZÁCIA LITERÁRNEJ SÚŤAŽE BÁSEŇ V ŠTÚROVČINE

S. Habijanec okrem nesprávne uvedenej informácie o organizátoroch súťaže spomína vo svojom článku aj vývesky v „štúrovčine“, ktoré mali údajne slúžiť ako inšpirácia potenciálnym záujemcom o súťaž (2023).

⁶ Niektoré ďalšie ahistorické lexikálne ekvivalenty sú alúziou na román Viliama Klimáčka *Nad a má čas*, ktorý (úmyselne?) používa faux štúrovskú slovenčinu.

Literárnu súťaž *Báseň v štúrovčine* neorganizovala skupina pracovníčok a pracovníkov ústavu – v roku 2022 organizovali literárnu súťaž *Báseň v štúrovčine* tri členky oddelenia Slovenského národného korpusu (Kristína Bobeková, Romana Krolčíková a Michaela Majerčíková) a v roku 2023 dve členky tohto oddelenia (Kristína Bobeková a Michaela Majerčíková).

Text propagačných materiálov – vývesiek – bol napísaný v „štúrovčine“ z čisto marketingových dôvodov, nie ako inšpirácia pre potenciálnych súťažiacich. Podľa nášho názoru sme siahli po modernom, atraktívnom a pútavom stvárnení propagačných materiálov aj v rámci textu – propagačné materiály boli zverejnené v „poštúrovčenej“ slovenčine. Samotný text propagačných materiálov bol úmyselne upravený. To značí, že sme sa nedržali výsledného prepisu, ktorý ponúka nástroj „ludevít“, ale vedome sme zasiahli do ponúknutého prepisu a „pritvrdili“ sme v mäkkosti v niektorých slovách.

Uvedomujeme si, že v štúrovčine by niektoré slová neboli zapísané spôsobom, aký sme zvolili, avšak bol to zámer. Vychádzali sme zo zaužívanej predstavy o prehnannej mäkkosti štúrovčiny medzi používateľmi jazyka. Úmyselne sme pracovali s „poštúrovčenou“ slovenčinou, aby sme zdôraznili jej mäkkosť oproti tvrdosti vo výslovnosti v niektorých oblastiach Slovenska. Pri spätnom pohľade pripúšťame, že zvolená taktika pri texte na propagačných materiáloch – použitie „poštúrovčenej“ slovenčiny – mohla vyvolať skreslenú predstavu o štúrovskej slovenčine, avšak takto zvolený text splnil svoj cieľ priam nadštandardne pri propagácii literárnej súťaže. Vďaka atraktívnejším marketingovým technikám sa nám podarilo zaujať širšiu verejnosť a priblížiť Jazykovedný ústav L. Štúra SAV, v. v. i., a jazyk používateľom aj inak, ako len cez pravidlá pravopisu, jazykové korpusy, jazykovú poradňu či slovníkový portál.

4. CIEĽ LITERÁRNEJ SÚŤAŽE BÁSEŇ V ŠTÚROVČINE

Hlavným cieľom literárnej súťaže bola podpora, resp. snaha motivovať deti, mládež i dospelých ku kreatívnemu písaniu, a vo všeobecnosti vzbudiť záujem o slovenský jazyk a jeho rôzne podoby, nie popularizácia nástroja „ludevít“. Ďalšími poňutkami na organizáciu literárnej súťaže bola nielen propagácia *Retrográdneho slovníka súčasnej slovenčiny. Slovné tvary na báze Slovenského národného korpusu* – ktorý v roku 2022 predstavoval knižnú odmenu pre prvé dve súťažné kategórie a pedagogičky s najväčším počtom prihlásených básní svojich žiakov a študentov, a v roku 2023 predstavoval knižnú odmenu pre prvú súťažnú kategóriu – ale aj propagácia špecializovaného oddelenia Slovenského národného korpusu zameriavajúceho sa na korpusovú lingvistiku a počítačové spracovanie prirodzeného jazyka, ktoré zastrešilo vecné odmeny pre výhercov vo všetkých kategóriách počas oboch ročníkov súťaže.

5. BÁSNE ŽIAKOV A ŠTUDENTOV AKO MATERIÁL NA VEDECKÚ ANALÝZU

Nie je najšťastnejšie – najmä z hľadiska vedeckej etiky – kriticky analyzovať texty žiakov a študentov. V druhom rade by bolo žiaduce overiť si, či je samotná kritika využitia nástroja na mieste. V článku totiž S. Habijanec uvádza ako fakt, že básne boli „v drvivej väčšine prípadov skoncipované pomocou aplikácie Ludevít“ (2023, s. 116 – 117). Podmienkou zaradenia do súťaže však nebolo použitie nástroja, podmienkou bolo napísať básne. Niektorí súťažiaci sa na nás obrátili s tým, že ovládajú ortografické zákonitosti štúrovčiny a rozhodli sa napísať básne na základe svojich vedomostí; iní súťažiaci sa zase tešili zo svojich básní a po prepísaní do štúrovčiny nadobudli pocit, že ich básne stratili *šmrnc*, tak si štúrovčinu upravili podľa seba; ďalší zase zasiahli do navrhovaných „prekladov“ konkrétnych slov, lebo sa im básne po použití nástroja „ludevít“ nerýmovali.

6. ZÁVER

S. Habijanec píše v doslove svojho článku, že „nepresnosť prekladača do štúrovskej slovenčiny zostala 16 rokov nepovšimnutá“ (2023, s. 121). Ale nepovšimnutá kým? K limitom nástroja „ludevít“ – viac či menej závažným – sa vyjadril už v roku 2006 samotný autor nástroja R. Garabík na konferencii *XVI. kolokvium mladých jazykovedcov* s príspevkom *O jednej skratke*, ktorej výstupom bol verejne dostupný zborník publikovaných príspevkov *Varia. XVI.* v roku 2009. O popularizačnom a zábavnom charaktere nástroja akiste svedčí aj forma uvedeného článku.

Autor kritiky takisto zavádzajúcim spôsobom argumentuje, že nástroj „ludevít“ bol „uvedený do prevádzky v roku 2015 pri príležitosti 150. výročia úmrtia Ľ. Štúra“⁷ (2023, s. 121). S. Habijancovi nemožno dať za pravdu, síce v roku 2015 organizoval Jazykovedný ústav Ľ. Štúra SAV, v. v. i., pod taktovkou L. Molnár Satinskej multižánrový festival *Štúr-žúr*, avšak pri príležitosti **200. výročia narodenia** Ľ. Štúra. Nástroj „ludevít“ nebol sprístupnený v roku 2015, ako tvrdí autor kritiky, ale bol sprístupnený verejnosti už v roku 2006^[6]. Možno oprávnené predpokladať, že počas *Roka Ľudovíta Štúra* a festivalu *Štúr-žúr* sa dostalo nástroju „ludevít“ výrazne väčšej pozornosti ako počas predchádzajúcich rokov a vzrástla jeho obľúbenosť u používateľov jazyka, i keď R. Garabík (2009) zachytáva neočakávanú popularitu nástroja už tesne po jeho zverejnení.

Takisto si dovoľíme polemizovať nad presvedčením S. Habijanca o tom, „či sa javí ako vhodné, aby táto nepresná a nespoľahlivá aplikácia bola umiestnená na webovej stránke vedeckovýskumnej inštitúcie“ (2023, s. 121). Nazdávame sa, že nástroj na oficiálnu webovú stránku Jazykovedného ústavu Ľudovíta Štúra SAV, v. v. i., patrí –

⁷ Základné faktografické údaje u autora kritiky pokrývajú. Dovoľujeme si upozorniť na výročný nesúlad (!) – rok 2015 bol vládou Slovenskej republiky vyhlásený za tzv. *Rok Ľudovíta Štúra*, v ktorom sme si pripomínali 200. výročie narodenia a **159. výročie úmrtia** Ľ. Štúra.

obzvlášť, keď je všeobecne známe, že strojový preklad obsahoval po celé desaťročia nepresnosti, neúplnosti, gramatické chyby v preklade a vyžaduje nevyhnutnú potrebu ľudskej kontroly a opráv.⁸

Autorský zámer pobaviť a popularizovať vedecký výskum jazyka sa stretáva u používateľov jazyka s kladnou odozvou. Na základe spätných reakcií od používateľov, nielen tohto nástroja, možno usudzovať, že verejnosť víta propagačno-popularizačné aktivity Jazykovedného ústavu Ľ. Štúra SAV, v. v. i., ktorými vychádza v ústrety používateľom jazyka v najrozmanitejších podobách. Osobnosť Ľ. Štúra, ako aj dobré meno váženej vedeckovýskumnej inštitúcie nesúcej jeho meno, nijkým spôsobom neutrpela existenciou nástroja „ludevít“, práve naopak. Netradičným a zároveň atraktívnym spôsobom sa nám podarilo vyvolať neočakávaný záujem⁹ o slovenský jazyk naprieč celým Slovenskom v rôznych vekových skupinách používateľov jazyka – nielen formou vytvorenia nástroja „ludevít“, ale aj vyhlásením dvoch ročníkov literárnej súťaže *Báseň v štúrovčine*.

Na záver by sme chceli zdôrazniť, že oddelenie Slovenského národného korpusu, ako aj autor nástroja „ludevít“ víta všetky relevantné kritické pripomienky, ktoré bude možné v budúcnosti využiť na prípadné zlepšenie nástroja a zverejnenie jeho ďalšej, presnejšej verzie.

Literatúra

- GARABÍK, R. (2009): O jednej skratke. In: Múscová, G. (ed.): *Varia. XVI. Zborník príspevkov zo XVI. kolokvia mladých jazykovedcov (Časť-Papiernička, 8. 11. – 10. 11. 2006)*. Bratislava: Slovenská jazykovedná spoločnosť pri SAV – Jazykovedný ústav Ľ. Štúra SAV, 152 – 158.
- HABIJANEC, S. (2023): Kritický náhľad na automatický prekladač do štúrovskej slovenčiny. In: *Slovenská reč*, 88/1, 113 – 122.
- SCHERRER, Y. – ERJAVEC, T. (2016): Modernising historical Slovene words. In: *Natural Language Engineering*, 22/6, 881 – 905. DOI: 10.1017/S1351324915000236
- ŠTÚR, L. (1846): *Nauka reči Slovenskej*. Prešporok: Tatrin.

Internetové zdroje

- [¹] https://metadata.ftp-master.debian.org/changelogs/main/l/ludevít/ludevít_9.2-1_changelog (cit. 15. 5. 2024).
- [²] <https://github.com/garabik/ludevít/commits/master/> (cit. 15. 5. 2024).
- [³] <https://www.euromatrixplus.net/> (cit. 15. 5. 2024).
- [⁴] <https://www.juls.savba.sk/glg/> (cit. 15. 5. 2024).
- [⁵] <https://kassiopeia.juls.savba.sk/~garabik/software/ludevít.html> (cit. 15. 5. 2024).
- [⁶] <https://korpus.sk/korpusy-a-databazy/nastroje-snk/ludevít/> (cit. 15. 5. 2024).

⁸ Čo prestáva platiť až teraz s nástupom používania neurónových sietí pri preklade a vznikom veľkých jazykových modelov.

⁹ V roku 2022 bolo do literárnej súťaže *Báseň v štúrovčine* prihlásených 150 básní z celého Slovenska a v roku 2023 až 456 básní.

Kristína Bobeková

Jazykovedný ústav Ľudovíta Štúra SAV, v. v. i.

Panská 26, 811 01 Bratislava

E-mail: kristina.bobekova@korpus.juls.savba.sk

Jana Levická

Jazykovedný ústav Ľudovíta Štúra SAV, v. v. i.

Panská 26, 811 01 Bratislava

E-mail: jana.levicka@korpus.juls.savba.sk

ORCID: 0000-0001-6027-604X

Michaela Majerčíková

Jazykovedný ústav Ľudovíta Štúra SAV, v. v. i.

Panská 26, 811 01 Bratislava

E-mail: michaela.majercikova@korpus.juls.savba.sk