An overview of Grammatical Error Correction for the twelve MultiGEC-2025 languages

Arianna Masciolini[1], Andrew Caines[2], Orphée De Clercq[3], Joni Kruijsbergen[3], Murathan Kurfalı[4], Ricardo Muñoz Sánchez[1], Elena Volodina[1], Robert Östling[5], Kais Allkivi[6], Špela Arhar Holdt[7], Ilze Auzina[8], Roberts Darģis[8], Elena Drakonaki[9], Jennifer-Carmen Frey[10], Isidora Glišić[11], Pinelopi Kikilintza[9], Lionel Nicolas[10], Mariana Romanyshyn[12], Alexandr Rosen[13], Alla Rozovskaya[14], Kristjan Suluste[15], Oleksiy Syvokon[16], Alexandros Tantos[9], Despoina-Ourania Touriki[9], Konstantinos Tsiotskas[9], Eleni Tsourilla[9], Vassilis Varsamopoulos[9], Katrin Wisniewski[17], Aleš Žagar[7], Torsten Zesch[18]

[1]Språkbanken Text, SFS, University of Gothenburg, Sweden;
[2]University of Cambridge, UK; [3]Ghent University, Belgium;
[4]RISE Research Institutes of Sweden, Sweden;
[5]Stockholm University, Sweden; [6]Tallinn University, Estonia;
[7]University of Ljubljana, Slovenia; [8]IMCS at the University of Latvia, Latvia;
[9]Aristotle University of Thessaloniki, Greece; [10]Eurac Research Bolzano, Italy;
[11]University of Iceland, Iceland; [12]Grammarly; [13]Charles University, Czech Republic;
[14]City University of New York (CUNY), USA; [15]Institute of the Estonian Language, Estonia;
[16]Microsoft; [17]Leipzig University, Germany;
[18]FernUniversität in Hagen, Germany

multigec@svenska.gu.se

Abstract

This overview is complementary to the comprehensive dataset description article for MultiGEC – a dataset for Multilingual Grammatical Error Correction including data for twelve European languages: Czech, English, Estonian, German, Greek, Icelandic, Italian, Latvian, Russian, Slovene, Swedish and Ukrainian.

It is well-known that in the field of Natural Language Processing (NLP) most publications tend to focus on the English language. While this is due to historical reasons (ease of publication, greater outreach, increased number of citations, etc.), it does leave other languages at a disadvantage across multiple tasks. The MultiGEC dataset was created as an attempt to counteract this effect. This report provides a historical overview of the evolution of GEC for each of the twelve languages in this dataset and provides a context for the work on the dataset and the related MultiGEC-2025 shared task.

1 Introduction

The task of Grammatical Error Correction (GEC) consists of rewriting texts in a given language to make them adhere to its norms. Despite the name of the task, corrections are not always strictly grammatical in nature, and may also aim to improve lexical choices, orthography and other aspects of language use.

As in many other areas of Natural Language Processing (NLP), there is a strong bias in GEC research towards English. The MultiGEC dataset (Masciolini et al., submitted) was created to stimulate GEC-related research in languages other than English and currently covers twelve European languages: Czech, English, Estonian, German, Greek, Icelandic, Italian, Latvian, Russian, Slovene, Swedish and Ukrainian.

Figure 1 shows the widely different number of GEC-related publications across the MultiGEC languages, and the very heavy bias towards English, which has an order of magnitude more publications than the runner-up. For generating the numbers in the graph, we searched the Association for Computational Linguistics (ACL) Anthology for

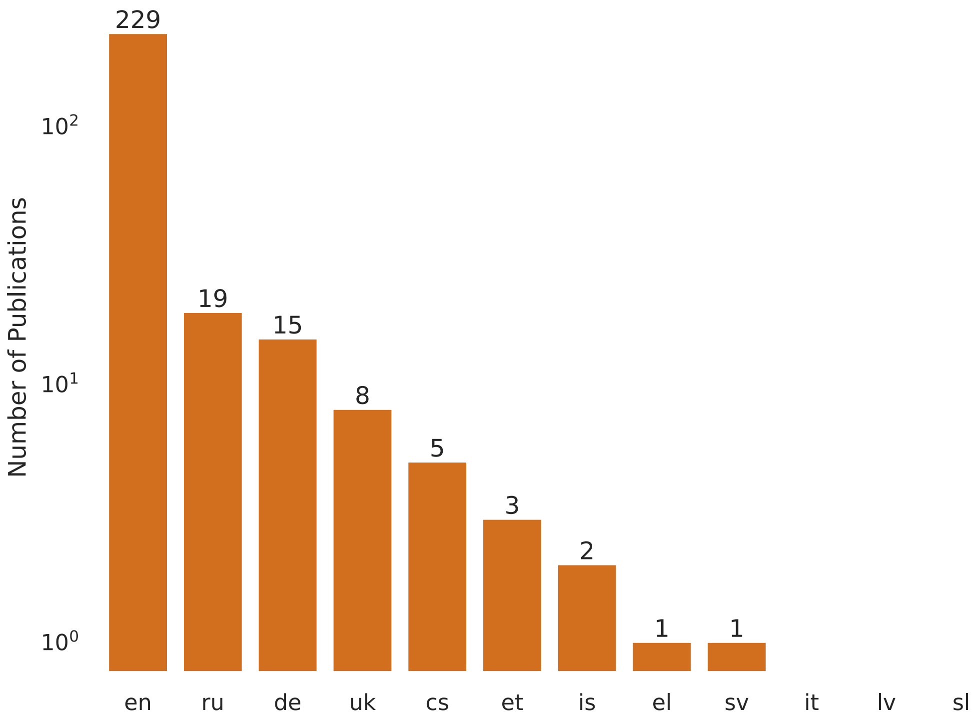

Figure 1: GEC-related publications in the ACL Anthology for the twelve MultiGEC languages.

co-occurrences of the terms "grammatical error correction"/"GEC" and the relevant language names in paper titles and abstracts, followed by a manual review of the hits. For publications that made no explicit mention of any specific language, we followed Ducel et al. (2022) and assumed English to be the object of study unless it was clearly stated that the work targeted multiple or low-resource languages.[1] This is an underestimation, as many articles (particularly older ones) are not indexed in the ACL Anthology, or were published before the term "grammatical error correction" was established. Our review cites a number of such articles, and all of the languages considered have at least a handful of articles.

This overview, complementary to the paper by Masciolini et al. (submitted) that introduces the original dataset, is meant to give more context for the need of such an initiative.

2 Overview of GEC Publications across the twelve MultiGEC Languages

Throughout this section we present an overview of the most prominent publications for the different MultiGEC languages in alphabetical order.

[1]The search was carried out on December 13, 2024.

2.1 Czech

Following the first spell checker for Czech (Hajič and Drózd, 1990), the earliest prototypes of a Czech grammar checker originated from a project using linguistically-motivated rules (Holan et al., 1997; Oliva, 1997). Rule-based methods are also behind two more recent systems: (Petkevič, 2014) and (Hlaváčková et al., 2022). A pre-trained model for Czech is also available for *Korektor* (Richter et al., 2012; Ramasamy et al., 2015), a statistical tool described as a "spellchecker and (occasional) grammar checker". Using neural Machine Translation (MT) approaches to GEC, Náplava and Straka (2019) achieved very competitive results for many languages, including Czech. A modified version of the system, trained on a large Czech corpus – content-wise identical to the Czech part of MultiGEC – scored even better (Náplava et al., 2022).

2.2 English

There have been many publications on GEC for English texts over the years. Interest in the topic accelerated rapidly in the early 2000s, with the development of learner-facing applications (Burstein et al., 2003) and research models targeting specific error types (e.g. De Felice and Pulman (2008) and Rozovskaya and Roth (2010)). Subsequently, the release of large annotated corpora and the organisation of shared tasks on English GEC gave added impetus and attention to the problem (e.f. Yannakoudakis et al., 2011; Ng et al., 2014; Bryant et al., 2019) and data-driven translation-based approaches became the dominant paradigm (e.g. Yuan and Briscoe, 2016; Grundkiewicz and Junczys-Dowmunt, 2018). More recently, researchers have been experimenting with prompting LLMs for GEC (Coyne et al., 2023; Fang et al., 2023b; Loem et al., 2023): for the time being, supervised models continue to be the state of the art, but LLMs may offer a different style of fluency edits compared to conventional miminal edit correction (Davis et al., 2024). It remains to be seen whether the latter will be (a) preferred by learners and (b) considered pedagogically useful. Bryant et al. (2023) offers a recent survey of GEC in English (and other languages).

2.3 Estonian

The first rule-based spell checker for Estonian was developed at the beginning of the 1990s,[2] followed by a rule-based tool for detecting errors in comma usage (Liin, 2009). More recently, statistical and neural methods have been applied to both spell-checking (Kaalep et al., 2022; Allkivi-Metsoja and Kippar, 2023) and sentence-level error correction. In the latter case, neural MT-based approaches are dominant (Korotkova et al., 2019; Luhtaru et al., 2024a). LLMs have recently been used for artificial error generation and correction (Luhtaru et al., 2024b). All in all, there is currently one Estonian GEC toolkit in active development.[3]

2.4 German

A wide range of methods for German has already been tested, especially on the Merlin corpus (Boyd et al., 2014). These include rule-based approaches (Kempfert and Köhn, 2018), LM-based re-ranking (Boyd, 2018), MT (Grundkiewicz and Junczys-Dowmunt, 2019; Katsumata and Komachi, 2020; Rothe et al., 2021; Luhtaru et al., 2024a) and even using multimodal data (Fang et al., 2023a).

2.5 Greek

There have been two notable efforts by Gakis et al. (2016) and Korre and Pavlopoulos (2022) to create GEC tools for Greek. The former resulted in a grammatical checker that achieves near-human accuracy for straightforward grammatical errors but does not address issues of cohesion, coherence and meaning. Korre and Pavlopoulos (2022) fine-tuned an MT5 multilingual text-to-text transformer using the Greek Native Corpus (GNC) (Korre et al., 2021) and evaluated its performance on both the GNC and the Greek Learner Corpus (GLCI) (Tantos and Papadopoulou, 2014). This work represents the most significant step toward enhancing GEC tools for Modern Greek.

2.6 Icelandic

The only open-source GEC tool for Icelandic, GreynirCorrect, uses the Icelandic Error Corpus (Ingason et al., 2021) to identify and correct grammar and spelling issues (Óladóttir et al., 2022). Today, two advanced neural GEC models exist for Icelandic: the byT5-base Transformer, a fine-tuned byte-level neural model (Ingólfsdóttir et al., 2024) and the GPT-SW3 model fine-tuned on Icelandic error corpora.

2.7 Italian

There are some (early) rule-based works on spelling correction that explicitly evaluate systems on Italian texts (Oflazer, 1996; Popescu and Vo, 2014; Gupta, 2020), next to systems for lexical normalization of social media texts (Weber and Zhekova, 2016; Van Der Goot et al., 2020). More recently, neural models for spelling correction have also been introduced (Sbattella and Tedesco, 2018; Ferrod et al., 2021). While the latter focuses exclusively on out-of-vocabulary words, the former operates at the sentence level and is potentially also capable of correcting grammatical errors.

2.8 Latvian

Latvian GEC is represented by several systems built using rule-based approaches and context free grammar formalisms, such as those described in Liin (2009), Deksne et al. (2014) and Deksne (2016); and recently through zero-shot neural monolingual MT approaches (Korotkova et al., 2019). In all these cases development of the systems has relied on error-annotated corpora of Latvian, not necessarily typical of language learners.

2.9 Russian

Similar to other languages, Russian GEC history can be traced back to approaches targeting

[2] Available as part of the Vabamorf morphological analysis toolkit: `github.com/Filosoft/vabamorf`

[3] `koodivaramu.eesti.ee/tartunlp/corrector`

spelling mistakes (Baytin, 2008; Panina et al., 2013; Sorokin et al., 2016; Sorokin, 2017; Rozovskaya, 2021). Beyond spell checking, GEC research has been facilitated by the release of the RULEC-GEC (Rozovskaya and Roth, 2019) and RU-Lang8 datasets (Trinh and Rozovskaya, 2021), resulting in a plethora of new models, such as neural MT approaches (Grundkiewicz and Junczys-Dowmunt, 2019; Náplava and Straka, 2019; Katsumata and Komachi, 2020; Kementchedjhieva and Søgaard, 2023) and LLMs (Flachs et al., 2021; Rothe et al., 2021; Sorokin, 2022; Katinskaia and Yangarber, 2023; Stahlberg and Kumar, 2024).

2.10 Slovene

There have been a few notable efforts in building and analyzing grammar-checking systems for Slovene, starting with a commercial grammar-checking tool (Holozan et al., 1992; Holozan, 2013). Recently, open-source tools have emerged. For instance, Petrič et al. (in press) applied SloBERTa to detect grammatical errors, achieving high accuracy on synthetic data but lower performance on real-world text, underscoring the need for authentic datasets. Božič et al. (2020) developed an automated comma placement tool. Finally, Klemen et al. (2024) introduced two Slovene spell-checkers: a traditional lexicon-based model and a neural model trained on synthetically generated errors.

2.11 Swedish

Swedish GEC history is split into two distinct periods: the first one in the 2000s, when development of rule-based (Birn, 2000) and statistical approaches to error detection and part-correction was ongoing within the Granska project (Domeij et al., 2000; Bigert and Knutsson, 2002); and the second period starting in 2021 with the release of the correction-annotated SweLL-gold corpus (Volodina et al., 2019), which led to the development of systems based on neural models, deep learning and LLMs (Nyberg, 2022; Kurfalı and Östling, 2023; Ehnroth and Park, 2023; Östling et al., 2024).

2.12 Ukranian

The first rule-based GEC system for Ukranian was contributed to LanguageTool by Andriy Rysin in 2007 and has been actively developed since then (cf. among others Miłkowski (2010)). Over the years, it has expanded to include extensive rule coverage and has facilitated the creation of various linguistic resources and tools, such as a part-of-speech tagger and the VESUM dictionary (Starko and Rysin, 2022). However, statistical methods for GEC remained underdeveloped. In 2021, the first corpus of Ukrainian GEC and fluency edits was released (Syvokon et al., 2023). This corpus served as the basis for a Shared Task on Ukrainian GEC (Syvokon and Romanyshyn, 2023), which stimulated the development of statistical models for Ukrainian GEC (Palma Gomez et al. (2023); Didenko and Sameliuk (2023); Bondarenko et al. (2023); Saini et al. (2024); *inter alia*).

3 Concluding remarks

We have given a short overview of the field of Grammatical Error Correction for the languages included in the MultiGEC corpus. Despite some work having been carried out for all 12 languages, comparison with English shows that the field would benefit from a more multilingual perspective, which in turn asks for more training and evaluation datasets for the less-represented languages. The latter problem is dual – on the one hand, it heavily depends on whether there has been any work on creating this sort of datasets for the concerned languages; on the other – it asks for an effort to promote such datasets in the community. This is the reason why we have compiled the MultiGEC dataset (Masciolini et al., submitted) – for pushing the field in a multilingual direction, and for promoting less-represented languages in the field of Grammatical Error Correction. The MultiGEC dataset has been used for the MultiGEC-2025 shared task and is available for new users.[4]

References

Kais Allkivi-Metsoja and Jaagup Kippar. 2023. Spelling Correction for Estonian Learner Language. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 782–788. University of Tartu Library.

Baytin. 2008. Search query correction in Yandex. In *Russian Internet technologies*.

Johnny Bigert and Ola Knutsson. 2002. Robust error detection: A hybrid approach combining unsupervised error detection and linguistic knowledge. In *Proc. 2nd Workshop Robust Methods in Analysis of Natural language Data (ROMAND'02), Frascati, Italy*, pages 10–19.

[4] github.com/spraakbanken/multigec-2025

Juhani Birn. 2000. Detecting grammar errors with Lingsoft's Swedish grammar checker. In *Proceedings of the 12th Nordic Conference of Computational Linguistics (NODALIDA 1999)*, pages 28–40.

Maksym Bondarenko, Artem Yushko, Andrii Shportko, and Andrii Fedorych. 2023. Comparative Study of Models Trained on Synthetic Data for Ukrainian Grammatical Error Correction. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 103–113. Association for Computational Linguistics.

Adriane Boyd. 2018. Using Wikipedia Edits in Low Resource Grammatical Error Correction. In *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, pages 79–84. Association for Computational Linguistics.

Adriane Boyd, Jirka Hana, Lionel Nicolas, Detmar Meurers, Katrin Wisniewski, Andrea Abel, Karin Schöne, Barbora Štindlová, and Chiara Vettori. 2014. The MERLIN corpus: Learner Language and the CEFR. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1281–1288. European Language Resources Association (ELRA).

Martin Božič, Marko Robnik Šikonja, Simon Krek, Špela Arhar Holdt, Iztok Kosem, Polona Gantar, Jaka Čibej, Kaja Dobrovoljc, Bojan Klemenc, and Cyprian Adam Laskowski. 2020. Commas 1.0: Automated Comma Placement Tool. [COBISS.SI-ID 66415363]. CJVT Tools.

Christopher Bryant, Mariano Felice, Øistein E. Andersen, and Ted Briscoe. 2019. The BEA-2019 Shared Task on Grammatical Error Correction. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 52–75. Association for Computational Linguistics.

Christopher Bryant, Zheng Yuan, Muhammad Reza Qorib, Hannan Cao, Hwee Tou Ng, and Ted Briscoe. 2023. Grammatical Error Correction: A Survey of the State of the Art. *Computational Linguistics*, pages 643–701.

Jill Burstein, Martin Chodorow, and Claudia Leacock. 2003. Criterion online essay evaluation: An application for automated evaluation of student essays. In *Proceedings of the Fifteenth Conference on Innovative Applications of Artificial Intelligence*.

Steven Coyne, Keisuke Sakaguchi, Diana Galvan-Sosa, Michael Zock, and Kentaro Inui. 2023. Analyzing the Performance of GPT-3.5 and GPT-4 in Grammatical Error Correction. *arXiv preprint arXiv:2303.14342*.

Christopher Davis, Andrew Caines, Ø Andersen, Shiva Taslimipoor, Helen Yannakoudakis, Zheng Yuan, Christopher Bryant, Marek Rei, and Paula Buttery. 2024. Prompting open-source and commercial language models for grammatical error correction of English learner text. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 11952–11967. Association for Computational Linguistics.

Rachele De Felice and Stephen G. Pulman. 2008. A Classifier-Based Approach to Preposition and Determiner Error Correction in L2 English. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 169–176. Coling 2008 Organizing Committee.

Daiga Deksne. 2016. A new phase in the development of a grammar checker for Latvian. In *Human Language Technologies–The Baltic Perspective*, pages 147–152. IOS Press.

Daiga Deksne, Inguna Skadiņa, and Raivis Skadiņš. 2014. Extended CFG formalism for grammar checker and parser development. In *Computational Linguistics and Intelligent Text Processing: 15th International Conference, CICLing 2014, Kathmandu, Nepal, April 6-12, 2014, Proceedings, Part I 15*, pages 237–249. Springer.

Bohdan Didenko and Andrii Sameliuk. 2023. RedPenNet for Grammatical Error Correction: Outputs to Tokens, Attentions to Spans. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 121–131. Association for Computational Linguistics.

Rickard Domeij, Ola Knutsson, Johan Carlberger, and Viggo Kann. 2000. Granska–an efficient hybrid system for Swedish grammar checking. In *Proceedings of the 12th Nordic Conference of Computational Linguistics (NODALIDA 1999)*, pages 49–56.

Fanny Ducel, Karën Fort, Gaël Lejeune, and Yves Lepage. 2022. Do we name the languages we study? the #BenderRule in LREC and ACL articles. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 564–573, Marseille, France. European Language Resources Association.

Joel Ehnroth and Yoonjoo Park. 2023. Correction of Grammatical Errors in Swedish. Master's thesis, Lund university.

Tao Fang, Jinpeng Hu, Derek F. Wong, Xiang Wan, Lidia S. Chao, and Tsung-Hui Chang. 2023a. Improving Grammatical Error Correction with Multimodal Feature Integration. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9328–9344. Association for Computational Linguistics.

Tao Fang, Shu Yang, Kaixin Lan, Derek F. Wong, Jinpeng Hu, Lidia S. Chao, and Yue Zhang. 2023b. Is ChatGPT a Highly Fluent Grammatical Error Correction System? A Comprehensive Evaluation. *arXiv preprint arXiv:2304.01746*.

Roger Ferrod, Enrico Brunetti, Luigi Di Caro, Chiara Di Francescomarino, Mauro Dragoni, Chiara Ghidini, Renata Marinello, Emilio Sulis, et al. 2021. A support for understanding medical notes: correcting spelling errors in Italian clinical records. In *CEUR*

WORKSHOP PROCEEDINGS, volume 3060, pages 19–28. CEUR-WS.

Simon Flachs, Felix Stahlberg, and Shankar Kumar. 2021. Data Strategies for Low-Resource Grammatical Error Correction. In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 117–122. Association for Computational Linguistics.

Panagiotis Gakis, Christos Panagiotakopoulos, Kyriakos Sgarbas, Christos Tsalidis, and Vassilios Verykios. 2016. Design and construction of the Greek grammar checker. *Digital Scholarship in the Humanities*, 32(3):554–576.

Roman Grundkiewicz and Marcin Junczys-Dowmunt. 2018. Near Human-Level Performance in Grammatical Error Correction with Hybrid Machine Translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 284–290. Association for Computational Linguistics.

Roman Grundkiewicz and Marcin Junczys-Dowmunt. 2019. Minimally-Augmented Grammatical Error Correction. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 357–363. Association for Computational Linguistics.

Prabhakar Gupta. 2020. A context-sensitive real-time Spell Checker with language adaptability. In *2020 IEEE 14th International Conference on Semantic Computing (ICSC)*, pages 116–122. IEEE.

Jan Hajič and Januš Drózd. 1990. Spelling-checking for Highly Inflective Languages. In *COLING 1990 Volume 3: Papers presented to the 13th International Conference on Computational Linguistics*, pages 358–360.

Dana Hlaváčková, Hana Žižková, Klára Dvořáková, and Markéta Pravdová. 2022. Developing Online Czech Proofreader Tool: Achievements, Limitations and Pitfalls. *Bohemistyka*, 21(1):122–134.

Tomáš Holan, Vladislav Kuboň, and Martin Plátek. 1997. A Prototype of a Grammar Checker for Czech. In *ANLC '97: Proceedings of the fifth conference on Applied natural language processing*, pages 147–154. Association for Computational Linguistics.

P. Holozan, M. Poznanovič, M. Pivec, and D. Sitar. 1992. *BesAna : besedna analiza : slovenski slovnični pregledovalnik. Navodila za uporabo*. Amebis.

Peter Holozan. 2013. Uporaba strojnega učenja za postavljanje vejic v slovenščini. *Uporabna informatika*, 21(4):196–209.

Anton Karl Ingason, Lilja Björk Stefánsdóttir, Þórunn Arnardóttir, and Xindan Xu. 2021. Icelandic Error Corpus (IceEC) Version 1.1. CLARIN-IS.

Svanhvít Lilja Ingólfsdóttir, Þórunn Arnardóttir, Pétur Orri Ragnarsson, Haukur Páll Jónsson, Haukur Barri Símonarson, Vilhjálmur Þorsteinsson, and Vésteinn Snæbjarnarson. 2024. Byte-Level Neural Error Correction Model for Icelandic - Yfirlestur (24.03). CLARIN-IS.

Heiki-Jaan Kaalep, Flammie Pirinen, and Sjur Nørstebø Moshagen. 2022. You can't suggest that?! Comparisons and improvements of speller error models. *Nordlyd. Morfologi, målstrev og maskinar: Trond Trosterud fyller | deavdá | turns 60!*, 46(1):125–139.

Anisia Katinskaia and Roman Yangarber. 2023. Grammatical Error Correction for Sentence-level Assessment in Language Learning. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 488–502. Association for Computational Linguistics.

Satoru Katsumata and Mamoru Komachi. 2020. Stronger Baselines for Grammatical Error Correction Using a Pretrained Encoder-Decoder Model. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 827–832. Association for Computational Linguistics.

Yova Kementchedjhieva and Anders Søgaard. 2023. Grammatical Error Correction through Round-Trip Machine Translation. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2208–2215. Association for Computational Linguistics.

Inga Kempfert and Christine Köhn. 2018. An Automatic Error Tagger for German. In *Proceedings of the 7th workshop on NLP for Computer Assisted Language Learning*, pages 32–40. LiU Electronic Press.

Matej Klemen, Martin Božič, Špela Arhar Holdt, and Marko Robnik-Šikonja. 2024. Neural Spell-Checker: Beyond Words with Synthetic Data Generation. In *International Conference on Text, Speech, and Dialogue*, pages 85–96. Springer.

Elizaveta Korotkova, Agnes Luhtaru, Maksym Del, Krista Liin, Daiga Deksne, and Mark Fishel. 2019. Grammatical Error Correction and Style Transfer via Zero-shot Monolingual Translation.

Katerina Korre, Marita Chatzipanagiotou, and John Pavlopoulos. 2021. ELERRANT: Automatic Grammatical Error Type Classification for Greek. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 708–717. INCOMA Ltd.

Katerina Korre and John Pavlopoulos. 2022. Enriching Grammatical Error Correction Resources for Modern Greek. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4984–4991. European Language Resources Association.

Murathan Kurfalı and Robert Östling. 2023. A distantly supervised Grammatical Error Detection/Correction system for Swedish. In *Proceedings of the 12th Workshop on NLP for Computer Assisted Language Learning (NLP4CALL)*.

Krista Liin. 2009. Grammar checker for detecting comma mistakes [Estonian]. *Philologia Estonica Tallinnensis*, 11:99–114.

Mengsay Loem, Masahiro Kaneko, Sho Takase, and Naoaki Okazaki. 2023. Exploring Effectiveness of GPT-3 in Grammatical Error Correction: A Study on Performance and Controllability in Prompt-Based Methods. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 205–219. Association for Computational Linguistics.

Agnes Luhtaru, Elizaveta Korotkova, and Mark Fishel. 2024a. No Error Left Behind: Multilingual Grammatical Error Correction with Pre-trained Translation Models. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1209–1222. Association for Computational Linguistics.

Agnes Luhtaru, Taido Purason, Martin Vainikko, Maksim Del, and Mark Fishel. 2024b. To Err Is Human, but Llamas Can Learn It Too. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. Association for Computational Linguistics.

Arianna Masciolini, Andrew Caines, Orphée De Clercq, Joni Kruijsbergen, Murathan Kurfalı, Ricardo Muñoz Sánchez, Elena Volodina, Robert Östling, Kais Allkivi-Metsoja, Špela Arhar Holdt, Ilze Auzina, Roberts Darģis, Elena Drakonaki, Jennifer-Carmen Frey, Isidora Glišić, Pinelopi Kikilintza, Lionel Nicolas, Mariana Romanyshyn, Alexandr Rosen, Alla Rozovskaya, Kristjan Suluste, Oleksiy Syvokon, Alexandros Tantos, Despoina-Ourania Touriki, Konstantinos Tsiotskas, Eleni Tsourilla, Vassilis Varsamopoulos, Katrin Wisniewski, Aleš Žagar, and Torsten Zesch. submitted. Towards better language representation – a multilingual dataset for text-level grammatical error correction. *International Journal of Learner Corpus Research*, TBA:TBA.

Marcin Miłkowski. 2010. Developing an open-source, rule-based proofreading tool. *Software: Practice and Experience*, 40(7):543–566.

Jakub Náplava and Milan Straka. 2019. Grammatical Error Correction in Low-Resource Scenarios. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 346–356. Association for Computational Linguistics.

Jakub Náplava, Milan Straka, Jana Straková, and Alexandr Rosen. 2022. Czech Grammar Error Correction with a Large and Diverse Corpus. *Transactions of the Association for Computational Linguistics*, 10:452–467.

Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The CoNLL-2014 Shared Task on Grammatical Error Correction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–14. Association for Computational Linguistics.

Martina Nyberg. 2022. *Grammatical Error Correction for Learners of Swedish as a Second Language*. Master's thesis, Uppsala university.

Jakub Náplava and Milan Straka. 2019. Grammatical Error Correction in Low-Resource Scenarios. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 346–356. Association for Computational Linguistics.

Kemal Oflazer. 1996. Error-tolerant Finite-state Recognition with Applications to Morphological Analysis and Spelling Correction. *Computational Linguistics*, 22(1):73–89.

Hulda Óladóttir, Þórunn Arnardóttir, Anton Karl Ingason, and Vilhjálmur Þorsteinsson. 2022. Developing a Spell and Grammar Checker for Icelandic Using an Error Corpus. In *2022 Language Resources and Evaluation Conference, LREC 2022*, 2022 Language Resources and Evaluation Conference, LREC 2022, pages 4644–4653. European Language Resources Association (ELRA).

Karel Oliva. 1997. Techniques for Accelerating a Grammar-Checker. In *Fifth Conference on Applied Natural Language Processing*, pages 155–158. Association for Computational Linguistics.

Robert Östling, Katarina Gillholm, Murathan Kurfalı, Marie Mattson, and Mats Wirén. 2024. Evaluation of Really Good Grammatical Error Correction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 6582–6593.

Frank Palma Gomez, Alla Rozovskaya, and Dan Roth. 2023. A Low-Resource Approach to the Grammatical Error Correction of Ukrainian. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 114–120. Association for Computational Linguistics.

Maria Panina, Alexey Baytin, and Irina Galinskaya. 2013. Context-independent autocorrection of query spelling errors. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2013"*.

Vladimír Petkevič. 2014. Kontrola české gramatiky. *SALi*, 5(2):48–86.

Timotej Petrič, Špela Arhar Holdt, and Marko Robnik-Šikonja. in press. Pomembnost realistične evalvacije: Primer popravkov sklona in števila v slovenščini z velikim jezikovnim modelom . *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*. Forthcoming.

Octavian Popescu and Ngoc Phuoc An Vo. 2014. Fast and accurate misspelling correction in large corpora. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1634–1642.

Loganathan Ramasamy, Alexandr Rosen, and Pavel Straňák. 2015. Improvements to Korektor: A Case Study with Native and Non-Native Czech. In *Proceedings ITAT 2015: Information Technologies – Applications and Theory*, volume 1422 of *CEUR Workshop Proceedings*, pages 73–80. CEUR-WS.org.

Michal Richter, Pavel Straňák, and Alexandr Rosen. 2012. Korektor – A System for Contextual Spell-Checking and Diacritics Completion. In *Proceedings of COLING 2012: Posters*, pages 1019–1028. The COLING 2012 Organizing Committee.

Sascha Rothe, Jonathan Mallinson, Eric Malmi, Sebastian Krause, and Aliaksei Severyn. 2021. A Simple Recipe for Multilingual Grammatical Error Correction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 702–707. Association for Computational Linguistics.

Alla Rozovskaya. 2021. Spelling Correction for Russian: A Comparative Study of Datasets and Methods. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1206–1216. INCOMA Ltd.

Alla Rozovskaya and Dan Roth. 2010. Generating Confusion Sets for Context-Sensitive Error Correction. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 961–970. Association for Computational Linguistics.

Alla Rozovskaya and Dan Roth. 2019. Grammar Error Correction in Morphologically Rich Languages: The Case of Russian. *Transactions of the Association for Computational Linguistics*, 7:1–17.

Aman Saini, Artem Chernodub, Vipul Raheja, and Vivek Kulkarni. 2024. Spivavtor: An Instruction Tuned Ukrainian Text Editing Model. In *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING 2024*, pages 95–108. ELRA and ICCL.

Licia Sbattella and Roberto Tedesco. 2018. How to simplify human-machine interaction: A text complexity calculator and a smart spelling corrector. In *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, pages 304–305.

Alexey Sorokin. 2017. Spelling Correction for Morphologically Rich Language: a Case Study of Russian. In *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, pages 45–53. Association for Computational Linguistics.

Alexey Sorokin. 2022. Improved grammatical error correction by ranking elementary edits. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 11416–11429. Association for Computational Linguistics.

Alexey Sorokin, Alexey Baytin, Irina Galinskaya, and Tatiana Shavrina. 2016. SpellRuEval: the first competition on automatic spelling correction for Russian. In *Proceedings of the International Conference "Dialogue"*.

Felix Stahlberg and Shankar Kumar. 2024. Synthetic Data Generation for Low-resource Grammatical Error Correction with Tagged Corruption Models. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 11–16. Association for Computational Linguistics.

Vasyl Starko and Andriy Rysin. 2022. VESUM: A Large Morphological Dictionary of Ukrainian As a Dynamic Tool. In *COLINS*, pages 61–70.

Oleksiy Syvokon, Olena Nahorna, Pavlo Kuchmiichuk, and Nastasiia Osidach. 2023. UA-GEC: Grammatical Error Correction and Fluency Corpus for the Ukrainian Language. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 96–102. Association for Computational Linguistics.

Oleksiy Syvokon and Mariana Romanyshyn. 2023. The UNLP 2023 Shared Task on Grammatical Error Correction for Ukrainian. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 132–137. Association for Computational Linguistics.

Alexandros Tantos and Despoina Papadopoulou. 2014. *Stand-off Annotation in Learner Corpora: Compiling the Greek Learner Corpus (GLC)*. Peter Lang Publications.

Viet Anh Trinh and Alla Rozovskaya. 2021. New Dataset and Strong Baselines for the Grammatical Error Correction of Russian. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4103–4111. Association for Computational Linguistics.

Rob Van Der Goot, Alan Ramponi, Tommaso Caselli, Michele Cafagna, and Lorenzo De Mattei. 2020. Norm It!: Lexical Normalization for Italian and Its Downstream Effects for Dependency Parsing. In *Proceedings of the 12th International Conference on Language Resources and Evaluation*, pages 6272–6278. European Language Resources Association (ELRA).

Elena Volodina, Lena Granstedt, Arild Matsson, Beáta Megyesi, Ildikó Pilán, Julia Prentice, Dan Rosén, Lisa Rudebeck, Carl-Johan Schenström, Gunlög Sundberg, et al. 2019. The SweLL language learner corpus: From design to annotation. *Northern European Journal of Language Technology (NEJLT)*, 6:67–104.

Daniel Weber and Desislava Zhekova. 2016. Tweetnorm: text normalization on Italian twitter data. In *Proceedings of the 13th Conference on Natural Language Processing*, pages 306–312.

Helen Yannakoudakis, Ted Briscoe, and Ben Medlock. 2011. A New Dataset and Method for Automatically Grading ESOL Texts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 180–189. Association for Computational Linguistics.

Zheng Yuan and Ted Briscoe. 2016. Grammatical error correction using neural machine translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 380–386. Association for Computational Linguistics.