A Crash Course in Automatic Grammatical Error Correction

Roman Grundkiewicz[†] and **Christopher Bryant**[‡] and **Mariano Felice**[‡]
[†] Microsoft / University of Edinburgh
[‡] ALTA Institute, University of Cambridge
rgrundki@inf.ed.ac.uk, {cjb255,mf501}@cl.cam.ac.uk

1 Motivation

Grammatical Error Correction (GEC) is the task of automatically detecting and correcting all types of errors in written text. Although most research has focused on correcting errors in the context of English as a Second Language (ESL) (Leacock et al., 2010), GEC can also be applied to other languages and native text. The main application of a GEC system is thus to assist humans with their writing.

Academic and commercial interest in GEC has grown significantly since the Helping Our Own (HOO) and Conference on Natural Language Learning (CoNLL) shared tasks in 2011-14 (Dale and Kilgarriff, 2011; Dale et al., 2012; Ng et al., 2013; Ng et al., 2014), and a record-breaking 24 teams took part in the recent Building Educational Applications (BEA) shared task (Bryant et al., 2019). Given this interest, and the recent shift towards neural approaches, we believe the time is right to offer a tutorial on GEC for researchers who may be new to the field or who are interested in the current state of the art and future challenges.

With this in mind, the main goal of this tutorial is not only to bring attendees up to speed with GEC in general, but also examine the development of neural-based GEC systems (Grundkiewicz et al., 2019; Choe et al., 2019a).

The tutorial will thus cover *introductory* material and near *cutting-edge* research.

2 Content

The tutorial will begin by introducing the task and describing some of its challenges in terms of the complexity of natural languages and technical limitations. For example, closed class errors, such as prepositions, are often easier to correct than open class errors, such as content words, because the confusion sets are smaller and they tend to require less semantic knowledge.

It will then continue by providing an overview of the evolution of approaches to the problem, including:

- Rule-based methods; e.g. Park et al. (1997), Schneider and McCoy (1998).
- Language model rescoring; e.g. Turner and Charniak (2007), Gamon et al. (2008).
- Error-type classifiers; e.g. De Felice and Pulman (2008), Tetreault et al. (2010), Rozovskaya and Roth (2011), Rozovskaya and Roth (2014).
- Statistical machine translation (SMT); e.g. Junczys-Dowmunt and Grundkiewicz (2016), Chollampatt and Ng (2017).
- Neural machine translation (NMT); e.g. Yuan and Briscoe (2016), Chollampatt and Ng (2018a), Junczys-Dowmunt et al. (2018), Zhao et al. (2019).

This will be followed by a review of GEC shared tasks for English, showing how they helped advance research in the field.

The next section will then introduce all English datasets available to GEC researchers, including the FCE (Yannakoudakis et al., 2011), Lang-8 (Mizumoto et al., 2011; Tajiri et al., 2012), NUCLE (Dahlmeier et al., 2013), JFLEG (Napoles et al., 2017) and W&I+LOCNESS (Bryant et al., 2019), and discuss challenges in annotation and corpus processing. It will also discuss standard evaluation procedures and metrics, including the M^2 Scorer (Dahlmeier and Ng, 2012), I-measure (Felice and Briscoe, 2015), GLEU

Proceedings of the 28th International Conference on Computational Linguistics: Tutorial Abstracts, pages 33–38
Barcelona, Spain (Online), December 12th, 2020.

(Napoles et al., 2016) and ERRANT (Bryant et al., 2017), and highlight their limitations (Grundkiewicz et al., 2015; Choshen and Abend, 2018; Chollampatt and Ng, 2018b).

After the break, the tutorial will continue by focusing on the development of state-of-the-art neural GEC systems (Grundkiewicz et al., 2019; Choe et al., 2019a). In particular, although SMT systems outperformed NMT systems for much of the past five years, a number of techniques, such as transfer learning from monolingual data, strong regularization, model ensembling, and using a large-scale language model, have closed this gap. We will also highlight the main challenges in neural GEC – data sparsity and sub-optimal decoding – and describe several proposed techniques to overcome them (Rei et al., 2017; Kasewa et al., 2018; Xie et al., 2018; Ge et al., 2018; Lichtarge et al., 2019; Grundkiewicz et al., 2019; Choe et al., 2019b; Chollampatt et al., 2019; Kiyono et al., 2019).

Finally, we will close the tutorial by discussing the findings of the BEA-2019 shared task (Bryant et al., 2019), and outlining future work and research directions. In particular, we will highlight the advent of low resource GEC, which strives to make it possible to build GEC systems for any language without human-annotated training data (cf. Bryant and Briscoe (2018), Stahlberg et al. (2019), Alikaniotis et al. (2019)).

3 Tutorial outline

A brief outline of the tutorial:

1. Introduction
   (a) Task definition
   (b) Linguistic and technical challenges
2. Historical approaches and shared tasks
   (a) Different approaches to error correction
   (b) Shared tasks on GEC
3. Data and evaluation
   (a) Annotation of error corpora
   (b) Artificial error generation
   (c) Evaluation procedures and metrics
   (d) Limitations of automatic and human evaluation
4. Neural GEC and current trends
   (a) GEC as a low-resource NMT task
   (b) Overcoming the data sparsity
   (c) Improving correction
   (d) Beyond the NMT approach
5. Recent and future work
   (a) Low resource GEC for other languages
   (b) Findings of the BEA-2019 shared task
   (c) Future directions

4 Prerequisites and reading list

A basic knowledge of machine learning (e.g. familiarity with precision, recall, F-score metrics, and ensemble methods) and neural approaches to natural language processing (e.g. the concept of sequence-to-sequence encoder-decoder neural models) will be helpful to understand the content of the tutorial.

Suggested reading list:

1. Automated grammatical error detection for language learners. Leacock, Chodorow, Gamon, and Tetreault (2010).

2. The CoNLL-2014 shared task on grammatical error correction. Ng, Wu, Briscoe, Hadiwinoto, Susanto, and Bryant (2014).

3. Building a state-of-the-art grammatical error correction system. Rozovskaya and Roth (2014).

4. Phrase-based machine translation is state-of-the-art for automatic grammatical error correction. Junczys-Dowmunt and Grundkiewicz (2016).

5. A multilayer convolutional encoder-decoder neural network for grammatical error correction. Chollampatt and Ng (2018).

6. Neural grammatical error correction systems with unsupervised pre-training on synthetic data. Grundkiewicz, Junczys-Dowmunt, and Heafield (2019).

7. An empirical study of incorporating pseudo data into grammatical error correction. Kiyono, Suzuki, Mita, Mizumoto, and Inui (2019).

5 Instructors

Roman Grundkiewicz is a Researcher at Microsoft Research and a Research Associate in the Institute for Language, Cognition and Computation at the University of Edinburgh. His research interests include machine translation and its applications to automatic grammatical error correction, automatic post-editing, and transliteration. He completed his PhD on "Algorithms for Automatic Grammatical Error Correction" at the Adam Mickiewicz University in Poznan, Poland. He participated in several shared tasks, winning the BEA shared task on GEC in 2019, and the WMT shared tasks on Automated Post-Editing in 2016 and 2018. His teaching experience includes courses on programming languages, artificial intelligence, natural language processing and machine learning. At the Machine Translation Marathons in 2018[1] and 2019[2] he ran tutorials on the Marian NMT toolkit[3] for an international audience.

Christopher Bryant[4] is a Research Associate in the ALTA Institute,[5] based in the Department of Computer Science and Technology at the University of Cambridge. His research interests include automatic grammatical error detection and correction, automatic corpus annotation, and computer aided language learning (CALL). He completed his PhD on "Automatic annotation for grammatical error correction" in 2019 and is the lead developer of the associated ERRor ANnotation Toolkit (ERRANT). Most recently, he also led the organisation of the BEA shared task on GEC in 2019, and has previously given a tutorial on GEC to an international audience at the first summer school on Machine Learning for Digital English Language Teaching in 2017.[6]

Mariano Felice[7] is a Research Associate in the ALTA Institute, based in the Department of Computer Science and Technology at the University of Cambridge. His research interests include grammatical error detection and correction, artificial error generation, automated generation of cloze tests and evaluation methods. He holds a PhD from the University of Cambridge, which focused on the generation of artificial errors to augment learner corpora for translation-based GEC. He was also one of the organisers of the BEA 2019 shared task on Grammatical Error Correction and a tutor of GEC at the first summer school on Machine Learning for Digital English Language Teaching in 2017.

[1] http://statmt.org/mtm18
[2] http://statmt.org/mtm19
[3] https://marian-nmt.github.io
[4] https://www.cl.cam.ac.uk/~cjb255/
[5] http://alta.cambridgeenglish.org/
[6] https://www.cl.cam.ac.uk/~hy260/ALTA-Summer-School-Chania-2017/
[7] https://www.cl.cam.ac.uk/~mf501/

References

Dimitris Alikaniotis, Vipul Raheja, and Joel Tetreault. 2019. The unreasonable effectiveness of transformer language models in grammatical error correction. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 127–133, Florence, Italy, August. Association for Computational Linguistics.

Christopher Bryant and Ted Briscoe. 2018. Language model based grammatical error correction without annotated training data. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 247–253, New Orleans, Louisiana, June. Association for Computational Linguistics.

Christopher Bryant, Mariano Felice, and Ted Briscoe. 2017. Automatic annotation and evaluation of error types for grammatical error correction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 793–805, Vancouver, Canada. Association for Computational Linguistics.

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, Florence, Italy, August. Association for Computational Linguistics.

Yo Joong Choe, Jiyeon Ham, Kyubyong Park, and Yeoil Yoon. 2019a. A neural grammatical error correction system built on better pre-training and sequential transfer learning. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 213–227, Florence, Italy, August. Association for Computational Linguistics.

Yo Joong Choe, Jiyeon Ham, Kyubyong Park, and Yeoil Yoon. 2019b. A neural grammatical error correction system built on better pre-training and sequential transfer learning. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 213–227, Florence, Italy, August. Association for Computational Linguistics.

Shamil Chollampatt and Hwee Tou Ng. 2017. Connecting the dots: towards human-level grammatical error correction. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 327–333, Copenhagen, Denmark, September. Association for Computational Linguistics.

Shamil Chollampatt and Hwee Tou Ng. 2018a. A multilayer convolutional encoder-decoder neural network for grammatical error correction. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, February.

Shamil Chollampatt and Hwee Tou Ng. 2018b. A reassessment of reference-based grammatical error correction metrics. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2730–2741, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Shamil Chollampatt, Weiqi Wang, and Hwee Tou Ng. 2019. Cross-sentence grammatical error correction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 435–445, Florence, Italy, July. Association for Computational Linguistics.

Leshem Choshen and Omri Abend. 2018. Automatic metric validation for grammatical error correction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1372–1382, Melbourne, Australia. Association for Computational Linguistics.

Daniel Dahlmeier and Hwee Tou Ng. 2012. Better evaluation for grammatical error correction. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 568–572. Association for Computational Linguistics.

Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. 2013. Building a large annotated corpus of learner English: The NUS corpus of learner English. In *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 22–31.

Robert Dale and Adam Kilgarriff. 2011. Helping our own: The HOO 2011 pilot shared task. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 242–249, Nancy, France, September. Association for Computational Linguistics.

Robert Dale, Ilya Anisimoff, and George Narroway. 2012. HOO 2012: A report on the preposition and determiner error correction shared task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 54–62, Montréal, Canada, June. 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 - Volume 1*, COLING '08, pages 169–176, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mariano Felice and Ted Briscoe. 2015. Towards a standard evaluation method for grammatical error detection and correction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 578–587, Denver, Colorado, May. Association for Computational Linguistics.

Michael Gamon, Jianfeng Gao, Chris Brockett, Alexandre Klementiev, William B Dolan, Dmitriy Belenko, and Lucy Vanderwende. 2008. Using contextual speller techniques and language modeling for ESL error correction. In *IJCNLP*, volume 8.

Tao Ge, Furu Wei, and Ming Zhou. 2018. Reaching human-level performance in automatic grammatical error correction: An empirical study. Technical report, Microsoft Research Technical Report.

Roman Grundkiewicz, Marcin Junczys-Dowmunt, and Edward Gillian. 2015. Human evaluation of grammatical error correction systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 461–470, Lisbon, Portugal, September. Association for Computational Linguistics.

Roman Grundkiewicz, Marcin Junczys-Dowmunt, and Kenneth Heafield. 2019. Neural grammatical error correction systems with unsupervised pre-training on synthetic data. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 252–263, Florence, Italy, August. Association for Computational Linguistics.

Marcin Junczys-Dowmunt and Roman Grundkiewicz. 2016. Phrase-based machine translation is state-of-the-art for automatic grammatical error correction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1546–1556, Austin, Texas, November. Association for Computational Linguistics.

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha, and Kenneth Heafield. 2018. Approaching neural grammatical error correction as a low-resource machine translation task. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 595–606, New Orleans, Louisiana, June. Association for Computational Linguistics.

Sudhanshu Kasewa, Pontus Stenetorp, and Sebastian Riedel. 2018. Wronging a right: Generating better errors to improve grammatical error detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4977–4983, Brussels, Belgium, October-November. Association for Computational Linguistics.

Shun Kiyono, Jun Suzuki, Masato Mita, Tomoya Mizumoto, and Kentaro Inui. 2019. An empirical study of incorporating pseudo data into grammatical error correction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1236–1242, Hong Kong, China, November. Association for Computational Linguistics.

Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault. 2010. Automated grammatical error detection for language learners. *Synthesis lectures on human language technologies*, 3(1):1–134.

Jared Lichtarge, Christopher Alberti, Shankar Kumar, Noam Shazeer, Niki Parmar, and Simon Tong. 2019. Corpora generation for grammatical error correction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, USA, June. Association for Computational Linguistics.

Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata, and Yuji Matsumoto. 2011. Mining revision log of language learning SNS for automated japanese error correction of second language learners. In *The 5th International Joint Conference on Natural Language Processing*, pages 147–155.

Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel Tetreault. 2016. GLEU without tuning. *arXiv preprint arXiv:1605.02592*.

Courtney Napoles, Keisuke Sakaguchi, and Joel Tetreault. 2017. JFLEG: A fluency corpus and benchmark for grammatical error correction. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 229–234, Valencia, Spain, April. Association for Computational Linguistics.

Hwee Tou Ng, Siew Mei Wu, Yuanbin Wu, Christian Hadiwinoto, and Joel Tetreault. 2013. The CoNLL-2013 shared task on grammatical error correction. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–12, Sofia, Bulgaria, August. Association for Computational Linguistics.

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, Baltimore, Maryland, June. Association for Computational Linguistics.

Jong C Park, Martha Stone Palmer, and Clay Washburn. 1997. An english grammar checker as a writing aid for students of english as a second language. In *ANLP*, page 24.

Marek Rei, Mariano Felice, Zheng Yuan, and Ted Briscoe. 2017. Artificial error generation with machine translation and syntactic patterns. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 287–292, Copenhagen, Denmark, September. Association for Computational Linguistics.

A. Rozovskaya and D. Roth. 2011. Algorithm selection and model adaptation for esl correction tasks. In *Association for Computational Linguistics*, Portland, Oregon, June. Association for Computational Linguistics.

A. Rozovskaya and D. Roth. 2014. Building a state-of-the-art grammatical error correction system. *Transactions of the Association for Computational Linguistics*.

David Schneider and Kathleen F. McCoy. 1998. Recognizing syntactic errors in the writing of second language learners. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2*, ACL '98, pages 1198–1204. Association for Computational Linguistics.

Felix Stahlberg, Christopher Bryant, and Bill Byrne. 2019. Neural grammatical error correction with finite state transducers. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, USA, June. Association for Computational Linguistics.

Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. 2012. Tense and aspect error correction for ESL learners using global context. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2*, ACL '12, pages 198–202, Stroudsburg, PA, USA. Association for Computational Linguistics.

Joel Tetreault, Jennifer Foster, and Martin Chodorow. 2010. Using parse features for preposition selection and error detection. In *Proceedings of the ACL 2010 Conference Short Papers*, ACLShort '10, pages 353–358, Stroudsburg, PA, USA. Association for Computational Linguistics.

Jenine Turner and Eugene Charniak. 2007. Language modeling for determiner selection. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, NAACL-Short '07, pages 177–180. Association for Computational Linguistics.

Ziang Xie, Guillaume Genthial, Stanley Xie, Andrew Ng, and Dan Jurafsky. 2018. Noising and denoising natural language: Diverse backtranslation for grammar correction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 619–628, New Orleans, Louisiana, June. Association for Computational Linguistics.

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, Portland, Oregon, USA, June. 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, San Diego, California, June. Association for Computational Linguistics.

Wei Zhao, Liang Wang, Kewei Shen, Ruoyu Jia, and Jingming Liu. 2019. Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 156–165, Minneapolis, Minnesota, June. Association for Computational Linguistics.