neural IR	# Semi-automated survey papers
Field:	Grammatical Error Correction (GEC)

1.Main Research Institutions and People

Research Institution	Affiliated Researchers
alibaba.com	Bibek Behera
baidu.com	Andrew Y Ng
cam.ac.uk	Christopher Bryant, Helen Yannakoudakis, Marek Rei, Mariano Felice, Oistein E Andersen, Ted Briscoe, Ekaterina Kochmar, Zheng Yuan
chungdahm.com	Soo-Hwa Lee,
cuny.edu	Alla Rozovskaya, Elena Filatova, Martin Chodorow
dcu.ie	Jennifer Foster
ed.ac.uk	Roman Grundkiewicz,
ets.org	Joel Tetreault, Nitin Madnani
grammarly.com	Claudia Leacock
harvard.edu	Keisuke Sakaguchi, Alexander Rush, Allen Schmaltz, Stuart Shieber, Yoon Kim
huji.ac.il	Leshem Choshen,Omri Abend
iit.ac.in	Pushpak Bhattacharyya
jhu.edu	Benjamin Van Durme, Courtney Napoles, Matt Post
kangwon.ac.kr	Jin-Young Ha
language-technology.com	Robert Dale
lexicalcomputing.com	Adam Kilgarriff
naist.jp	Ippei Yoshimoto, Tomoya Kose, Yuji Matsumoto, Yuta Hayashibe, Kensuke Mitsuzawa
nus.edu.sg	Kaveh Taghipour, Raymond Hendy Susanto, Shamil Chollampatt, Christian Hadiwinoto, Duc Tam Hoang, Hwee Tou Ng, Peter Phandi, Siew Mei Wu
nyu.edu	Nizar Habash
ox.ac.uk	Stephen G Pulman
pitt.edu	Na-Rae Han
pku.edu.cn	Tao Ge
research.google.com	Kristina Toutanova
research.microsoft.com	Chris Brockett, Furu Wei, Jianfeng Gao, Jianshu Ji, Marcin Junczys-Dowmunt, Michael Gamon, Ming Zhou, Qinlong Wang, Steven Truong, William B Dolan Yongen Gong
riken.jp	Hiroki Asano, Kentaro Inui, Tomoya Mizumoto
sap.com	Daniel Dahlmeier
stanford.edu	Anand Avati, Dan Jurafsky, Naveen Arivazhagan, Ziang Xie
tmu.ac.jp	Mamoru Komachi, Masahiro Kaneko, Yuya Sakaizawa
ucl.ac.uk	Rachele De Felice
ucla.edu	Kai-Wei Chang
uiuc.edu	Dan Roth, Mark Sammons

2.Motivation of the field of GEC

Motivation phrases	Paper ID
One of the fastest growing areas in need of NLP tools is the field of grammatical error correction for learners of English as a Second Language	[P11-2089]
It has been estimated that over a billion people are using or learning English as a second or foreign language, and the numbers are growing not only for English but for other languages as well. These language learners provide a burgeoning market for tools that help identify and correct learners' writing errors. Unfortunately, the errors targeted by typical commercial proofreading tools do not include those aspects of a second language that are hardest to learn.	[AutomatedGram]
Automated error detection for language learners is an underexplored area for state-of-the- art NLP research that can potentially benefit the very large and under-served community of language learners.	[AutomatedGram]
A system that performs poorly overall may in fact outperform others at specific error types. This is significant because a robust specialised system is actually more desirable than a mediocre general system	[P17-1074]
Tackling this task has far-reaching impact, since it is estimated that hundreds of millions of people worldwide are learning English and they benefit directly from an automated grammar checker.	[W14-1701]
One of the major advantages of using SMT in ESL writing assistance is that it can be expected to benefit automatically from any progress made in SMT itself .	[P06-1032]
Some substantially-sized annotated error corpora do exist, e.g. the Cambridge Learner Corpus, but these are not freely available. One way around this problem of lack of availability of suitably large error-annotated corpora is to introduce errors into sentences automatically	[Foster-BEA4]
A GEC system to correct English promises to benefit millions of learners around the world, since it functions as a learning aid by providing instantaneous feedback on ESL writing.	[D14-1102]
The SMT framework largely benefits from its ability to incorporate large error-corrected parallel corpora like the publicly available Lang-8 corpus (Mizumoto et al. 2011), additional English corpora for training robust language models (LMs), task-specific features (Junczys-Dowmunt and Grundkiewicz 2016), and neural models (Chollampatt, Taghipour, and Ng 2016). However, SMT-based systems suffer from limited generalization capabilities compared to neural approaches and are unable to access longer source and target contexts effectively.	[1801.08831]
Substantial performance improvements were achieved by training the best model on additional datasets. We found that the largest benefit was obtained from training on 8 million tokens of text from learners with varying levels of language proficiency.	[1607.06153]
Grammatical error correction (GEC) systems strive to correct both global errors in word order and usage, and local errors in spelling and inflection.	[1707.02026]
Shortage of available training data is holding back progress in the area of automated error detection.	[1707.05236]
Our analysis shows that the created systems are closer to reaching human-level performance than any other GEC system reported so far.	[1804.05945]
We consider grammar correction as a translation problem - translation from an incorrect sentence to a correct sentence. The correcting system is trained using a parallel corpus of incorrect and their corresponding correct sentences.	[I13-1122]
Progress in natural language processing (NLP) research is driven and measured by automatic evaluation methods. Automatic evaluation allows fast and inexpensive feedback during development, and objective and reproducible evaluation during testing time.	[N12-1067]

3.Problems

Problem Statements	Paper Id
Most approaches to error correction for non-native text are based on machine learning classifiers for specific error types.	[W13-3607]
How do we know which grammatical error correction (GEC) system is best? A number of metrics have been proposed over the years, each motivated by weaknesses of previous metrics; however, the metrics themselves have not been compared to an empirical gold standard grounded in human judgments.	[P15-2097]
Despite growing demand for ESL proofing tools, there has been remarkably little progress in this area over the last decade. Research into computer feedback for ESL writers remains largely focused on smallscale pedagogical systems implemented within the framework of CALL (Computer Aided Language Learning) (Reuer 2003; Vanderventer Faltin, 2003), while commercial ESL grammar checkers remain brittle and difficult to customize to meet the needs of ESL writers of different first-language (L1) backgrounds and skill levels.	[P06-1032]
One can argue that traditional understanding of grammar and grammar correction encompasses the idea of native-language fluency. However, the metrics commonly used in evaluating GEC undermine these arguments. The performance of GEC systems is typically evaluated using metrics that compute corrections against error-coded corpora, which impose a taxonomy of types of grammatical errors. Assigning these codes can be difficult, as evidenced by the low agreement found between annotators of these corpora. It is also quite expensive. But most importantly, as we will show in this paper, annotating for explicit error codes places a downward pressure on annotators to find and fix concrete, easily-identifiable grammatical errors (such as wrong verb tense) in lieu of addressing the native fluency of the text.	[Q16-1013]
Metric validation in Grammatical Correction (GEC) is currently done observing the correlation between human and metric-induced rankings. However, such correlation studies are methodologically troublesome, and from low inter-rater agreement.	[1804.11225]
Automatically generated ungrammatical data has been used in the training and evaluation of error detection systems. The main advantage of using such artificial data is that it is cheap to produce. However, it is of little use if it is not a realistic model of the naturally occurring and expensive data that it is designed to replace.	[Foster-BEA4]
Modeling only the context of the words from a raw corpus written by native speakers does not consider specific grammatical errors of language learners. This leads to the problem wherein the word embeddings of correct and incorrect expressions tend to be similar so that the classifier must decide grammaticality of a word from contextual information with a similar input vector.	[I17-1005]
The GEC mainly constitutes a generative task, i.e., a task that produces a grammatically correct sentence from a given original sentence whereby multiple distinct outputs can be judged “correct” for a single input. Therefore, automatically evaluating the performance is not straightforward and is considered as an important issue as in the fields of translation and summarization.	[I17-2058]

4.Approach

Main Approach	Paper Id
Classification	[W07-1604],[C08-1022],[han-lrec10-final],[AutomatedGram],[P10-2065],[W11-2838],[W14-1703],[P16-1208]
SMT	[P06-1032],[D11-1010], [N12-1067], [W13-3604], [I13-1122], [W14-1702], [W14-1704], [W14-1703],[P16-1208]
NMT	[N16-1042], [1603.09727], [1707.02026], [1707.00299], [1707.09067], [1801.08831],[1804.05945], [1804.05940], [1807.01270]
New	[D14-1102], [1606.00189],[D16-1195], [P16-1208], [1605.06353], [N16-1133], [W16-0530], [1606.00210], [D17-1297]
Evaluation	[P06-1032], [Foster-BEA4], [N10-1018], [P11-1093], [W12-2032], [E14-3013], [1603.09727], [1707.05236],[1603.09727]

5.Abstracts

Paper ID	Abstract
N10-1018	This paper proposes a novel approach to the problem of training classifiers to detect and correct grammar and usage errors in text by selectively introducing mistakes into the training data. When training a classifier, we would like the distribution of examples seen in training to be as similar as possible to the one seen in testing. In error correction problems, such as correcting mistakes made by second language learners, a system is generally trained on correct data, since annotating data for training is expensive. Error generation methods avoid expensive data annotation and create training data that resemble non-native data with errors. We apply error generation methods and train classifiers for detecting and correcting article errors in essays written by non-native English speakers; we show that training on data that contain errors produces higher accuracy when compared to a system that is trained on clean native data. We propose several training paradigms with error generation and show that each such paradigm is superior to training a classifier on native data. We also show that the most successful error generation methods are those that use knowledge about the article distribution and error patterns observed in non-native text.
P11-1093	We consider the problem of correcting errors made by English as a Second Language (ESL) writers and address two issues that are essential to making progress in ESL error correction - algorithm selection and model adaptation to the first language of the ESL learner A variety of learning algorithms have been applied to correct ESL mistakes, but often comparisons were made between incomparable data sets. We conduct an extensive, fair comparison of four popular learning methods for the task, reversing conclusions from earlier evaluations. Our results hold for different training sets, genres, and feature sets; A second key issue in ESL error correction is the adaptation of a model to the first language of the writer. Errors made by non-native speakers exhibit certain regularities and, as we show, models perform much better when they use knowledge about error patterns of the nonnative writers. We propose a novel way to adapt a learned algorithm to the first language of the writer that is both cheaper to implement and performs better than other adaptation methods
P16-1208	We focus on two leading state-of-the-art approaches to grammatical error correction machine learning classification and machine translation. Based on the comparative study of the two learning frameworks and through error analysis of the output of the state-of-the-art systems, we identify key strengths and weaknesses of each of these approaches and demonstrate their complementarity. In particular, the machine translation method learns from parallel data without requiring further linguistic input and is better at correcting complex mistakes. The classification approach possesses other desirable characteristics, such as the ability to easily generalize beyond what was seen in training, the ability to train without human-annotated data, and the flexibility to adjust knowledge sources for individual error types. Based on this analysis, we develop an algorithmic approach that combines the strengths of both methods. We present several systems based on resources used in previous work with a relative improvement of

	over 20% (and 7.4 F score points) over the previous state-of-the-art.
W14-1704	The CoNLL-2014 shared task is an extension of last year's shared task and focuses on correcting grammatical errors in essays written by non-native learners of English. In this paper, we describe the Illinois-Columbia system that participated in the shared task. Our system ranked second on the original annotations and first on the revised annotations. The core of the system is based on the University of Illinois model that placed first in the CoNLL-2013 shared task. This baseline model has been improved and expanded for this year's competition in several respects. We describe our underlying approach, which relates to our previous work, and describe the novel aspects of the system in more detail.
W12-2032	We describe the University of Illinois (UI) system that participated in the Helping Our Own (HOO) 2012 shared task, which focuses on correcting preposition and determiner errors made by non-native English speakers. The task consisted of three metrics: Detection, Recognition, and Correction, and measured performance before and after additional revisions to the test data were made. Out of 14 teams that participated, our system scored first in Detection and Recognition and second in Correction before the revisions; and first in Detection and second in the other metrics after revisions. We describe our underlying approach, which relates to our previous work in this area, and propose an improvement to the earlier method, error inflation, which results in significant gains in performance.
1707.090 67	In a controlled experiment of sequence-to-sequence approaches for the task of sentence correction, we find that character-based models are generally more effective than word-based models and models that encode sub-word information via convolutions, and that modeling the output data as a series of diffs improves effectiveness over standard approaches. Our strongest sequence-to-sequence model improves over our strongest phrase-based statistical machine translation model, with access to the same data, by 6 M2 (0.5 GLEU) points. Additionally, in the data environment of the standard CoNLL-2014 setup, we demonstrate that modeling (and tuning against) diffs yields similar or better M2 scores with simpler models and/or significantly less data than previous sequence-to-sequence approaches.
I13-1122	We introduce a novel technique that uses hierarchical phrase-based statistical machine translation (SMT) for grammar correction. SMT systems provide a uniform platform for any sequence transformation task. Thus grammar correction can be considered a translation problem from incorrect text to correct text. Over the years, grammar correction data in the electronic form (i.e., parallel corpora of incorrect and correct sentences) has increased manifolds in quality and quantity, making SMT systems feasible for grammar correction. Firstly, sophisticated translation models like hierarchical phrase-based SMT can handle errors as complicated as reordering or insertion, which were difficult to deal with previously through the mediation of rule based systems. Secondly, this SMT based correction technique is similar in spirit to human correction, because the system extracts grammar rules from the corpus and later uses these rules to translate incorrect sentences to correct sentences. We describe how to use Joshua, a hierarchical

	phrase-based SMT system for grammar correction. An accuracy of 0.77 (BLEU score) establishes the efficacy of our approach.
P06-1032	This paper presents a pilot study of the use of phrasal Statistical Machine Translation (SMT) techniques to identify and correct writing errors made by learners of English as a Second Language (ESL). Using examples of mass noun errors found in the Chinese Learner Error Corpus (CLEC) to guide creation of an engineered training set, we show that application of the SMT paradigm can capture errors not well addressed by widely-used proofing tools designed for native speakers. Our system was able to correct 61.81% of mistakes in a set of naturally occurring examples of mass noun errors found on the World Wide Web, suggesting that efforts to collect alignable corpora of pre- and post-editing ESL writing samples offer can enable the development of SMT-based writing assistance tools capable of repairing many of the complex syntactic and lexical problems found in the writing of ESL learners.
P17-1074	Until now, error type performance for Grammatical Error Correction (GEC) systems could only be measured in terms of recall because system output is not annotated. To overcome this problem, we introduce ERRANT, a grammatical ERRor ANnotation Toolkit designed to automatically extract edits from parallel original and corrected sentences and classify them according to a new, dataset-agnostic, rule-based framework. This not only facilitates error type evaluation at different levels of granularity, but can also be used to reduce annotator workload and standardise existing GEC datasets. Human experts rated the automatic edits as Good or Acceptable in at least 95% of cases, so we applied ERRANT to the system output of the CoNLL-2014 shared task to carry out a detailed error type analysis for the first time.
AutomatedGram	It has been estimated that over a billion people are using or learning English as a second or foreign language, and the numbers are growing not only for English but for other languages as well. These language learners provide a burgeoning market for tools that help identify and correct learners' writing errors. Unfortunately, the errors targeted by typical commercial proofreading tools do not include those aspects of a second language that are hardest to learn. This volume describes the types of constructions English language learners find most difficult - constructions containing prepositions, articles, and collocations. It provides an overview of the automated approaches that have been developed to identify and correct these and other classes of learner errors in a number of languages. Error annotation and system evaluation are particularly important topics in grammatical error detection because there are no commonly accepted standards. Chapters in the book describe the options available to researchers, recommend best practices for reporting results, and present annotation and evaluation schemes.
1610.02124	Current methods for automatically evaluating grammatical error correction (GEC) systems rely on gold-standard references. However, these methods suffer from penalizing grammatical edits that are correct but not in the gold standard. We show that reference-less grammaticality metrics correlate very strongly with human judgments and are competitive with the leading reference-based evaluation metrics. By interpolating both methods, we achieve state-of-the-art correlation with human judgments. Finally, we show that GEC metrics are much more reliable when they are calculated at the

	sentence level instead of the corpus level. We have set up a CodaLab site for benchmarking GEC output using a common dataset and different evaluation metrics.
1702.04066	We present a new parallel corpus, JHU FLuency-Extended GUG corpus (JFLEG) for developing and evaluating grammatical error correction (GEC). Unlike other corpora, it represents a broad range of language proficiency levels and uses holistic fluency edits to not only correct grammatical errors but also make the original text more native sounding. We describe the types of corrections made and benchmark four leadingGEC systems on this corpus, identifying specific areas in which they do well and how they can improve. JFLEG fulfills the need for a new gold standard to properly assess the current state ofGEC.
P15-2097	How do we know which grammatical error correction (GEC) system is best? A number of metrics have been proposed over the years, each motivated by weaknesses of previous metrics; however, the metrics themselves have not been compared to an empirical gold standard grounded in human judgments. We conducted the first human evaluation of GEC system outputs, and show that the rankings produced by metrics such as MaxMatch and I-measure do not correlate well with this ground truth. As a step towards better metrics, we also propose GLEU, a simple variant of BLEU, modified to account for both the source and the reference, and show that it hews much more closely to human judgments.
D11-1010	We present a novel approach for automatic collocation error correction in learner English which is based on paraphrases extracted from parallel corpora. Our key assumption is collocation errors are often caused by semantic similarity in the first language (L1 language) of the writer. An analysis of a large corpus of annotated learner English confirms this assumption. We evaluate our approach on real-world learner data and show that L1-induced paraphrases outperform traditional approaches based on edit distance, homophones, and WordNet synonyms.
D12-1052	We present a novel beam-search decoder for grammatical error correction. The decoder iteratively generates new hypothesis corrections from current hypotheses and scores them based on features of grammatical correctness and fluency. These features include scores from discriminative classifiers for specific error categories, such as articles and prepositions. Unlike all previous approaches, our method is able to perform correction of whole sentences with multiple and interacting errors while still taking advantage of powerful existing classifier approaches. Our decoder achieves an F1 correction score significantly higher than all previous published scores on the Helping Our Own (HOO) shared task data set.
N12-1067	We present a novel method for evaluating grammatical error correction. The core of our method, which we call MaxMatch (M2), is an algorithm for efficiently computing the sequence of phrase-level edits between a source sentence and a system hypothesis that achieves the highest overlap with the goldstandard annotation. This optimal edit sequence is subsequently scored using F1 measure. We test our M2 scorer on the Helping Our Own (HOO) shared task data and show that our method results in more accurate evaluation for grammatical error correction.
D17-1297	We propose an approach to N-best list reranking using neural sequence-labelling models. We train a compositional model for

	error detection that calculates the probability of each token in a sentence being correct or incorrect, utilising the full sentence as context. Using the error detection model, we then re-rank the N best hypotheses generated by statistical machine translation systems. Our approach achieves state-of-the-art results on error correction for three different datasets, and it has the additional advantage of only using a small set of easily computed features that require no linguistic input.
I17-2058	In grammatical error correction (GEC), automatically evaluating system outputs requires gold-standard references, which must be created manually and thus tend to be both expensive and limited in coverage. To address this problem, a referenceless approach has recently emerged; however, previous reference-less metrics that only consider the criterion of grammaticality, have not worked as well as reference-based metrics. This study explores the potential of extending a prior grammaticality-based method to establish a reference-less evaluation method for GEC systems. Further, we empirically show that a reference-less metric that combines fluency and meaning preservation with grammaticality provides a better estimate of manual scores than that of commonly used reference-based metrics. To our knowledge, this is the first study that provides empirical evidence that a reference-less metric can replace reference-based metrics in evaluating GEC systems.
W14-1701	The CoNLL-2014 shared task was devoted to grammatical error correction of all error types. In this paper, we give the task definition, present the data sets, and describe the evaluation metric and scorer used in the shared task. We also give an overview of the various approaches adopted by the participating teams, and present the evaluation results. Compared to the CoNLL2013 shared task, we have introduced the following changes in CoNLL-2014: (1) A participating system is expected to detect and correct grammatical errors of all types, instead of just the five error types in CoNLL-2013; (2) The evaluation metric was changed from F1 to F0.5, to emphasize precision over recall; and (3) We have two human annotators who independently annotated the test essays, compared to just one human annotator in CoNLL-2013.
W13-3604	This paper describes the Nara Institute of Science and Technology (NAIST) error correction system in the CoNLL 2013 Shared Task. We constructed three systems: a system based on the Treelet Language Model for verb form and subjectverb agreement errors; a classifier trained on both learner and native corpora for noun number errors; a statistical machine translation (SMT)-based model for preposition and determiner errors. As for subject-verb agreement errors, we show that the Treelet Language Model-based approach can correct errors in which the target verb is distant from its subject. Our system ranked fourth on the official run.
Foster-BEA4	This paper explores the issue of automatically generated ungrammatical data and its use in error detection, with a focus on the task of classifying a sentence as grammatical or ungrammatical. We present an error generation tool called GenERRate and show how Gen- ERRate can be used to improve the performance of a classifier on learner data. We describe initial attempts to replicate Cambridge Learner Corpus errors using GenERRate.
1707.02026	Grammatical error correction (GEC) systems strive to correct both global errors in word order and usage, and local errors in spelling and inflection. Further developing upon recent work on

	neural machine translation, we propose a new hybrid neural model with nested attention layers for GEC. Experiments show that the new model can effectively correct errors of both types by incorporating word and character-level information, and that the model significantly outperforms previous neural models for GEC as measured on the standard CoNLL14 benchmark dataset. Further analysis also shows that the superiority of the proposed model can be largely attributed to the use of the nested attention mechanism, which has proven particularly effective in correcting local errors that involve small edits in orthography.
P10-2065	We evaluate the effect of adding parse features to a leading model of preposition usage. Results show a significant improvement in the preposition selection task on native speaker text and a modest increment in precision and recall in an ESL error detection task. Analysis of the parser output indicates that it is robust enough in the face of noisy non-native writing to extract useful information.
W10-1006	In this paper we present results from two pilot studies which show that using the Amazon Mechanical Turk for preposition error annotation is as effective as using trained raters, but at a fraction of the time and cost. Based on these results, we propose a new evaluation method which makes it feasible to compare two error detection systems tested on different learner data sets.
Q16-1013	The field of grammatical error correction (GEC) has grown substantially in recent years, with research directed at both evaluation metrics and improved system performance against those metrics. One unvisited assumption, however, is the reliance of GEC evaluation on error-coded corpora, which contain specific labeled corrections. We examine current practices and show that GEC's reliance on such corpora unnaturally constrains annotation and automatic evaluation, resulting in (a) sentences that do not sound acceptable to native speakers and (b) system rankings that do not correlate with human judgments. In light of this, we propose an alternate approach that jettisons costly error coding in favor of unannotated, whole-sentence rewrites. We compare the performance of existing metrics over different gold-standard annotations, and show that automatic evaluation with our new annotation scheme has very strong correlation with expert rankings (. = 0.82). As a result, we advocate for a fundamental and necessary shift in the goal of GEC, from correcting small, labeled error types, to producing text that has native fluency.
1707.00299	We propose a neural encoder-decoder model with reinforcement learning (NRL) for grammatical error correction (GEC). Unlike conventional maximum likelihood estimation (MLE), the model directly optimizes towards an objective that considers a sentence-level, task-specific evaluation metric, avoiding the exposure bias issue in MLE. We demonstrate that NRL outperforms MLE both in human and automated evaluation metrics, achieving the state-of-the-art on a fluency-oriented GEC corpus.
W14-1703	Statistical machine translation toolkits like Moses have not been designed with grammatical error correction in mind. In order to achieve competitive results in this area, it is not enough to simply add more data. Optimization procedures need to be customized, task-specific features should be introduced. Only then can the decoder take advantage of relevant data. We demonstrate the validity of the above claims by combining web-scale language

	models and large-scale error-corrected texts with parameter tuning according to the task metric and correction-specific features. Our system achieves a result of 35.0% F0.5 on the blind CoNLL-2014 test set, ranking on third place. A similar system, equipped with identical models but without tuned parameters and specialized features, stagnates at 25.4%.
1605.06353	In this work, we study parameter tuning towards the M2 metric, the standard metric for automatic grammar error correction (GEC) tasks. After implementing M2 as a scorer in the Moses tuning framework, we investigate interactions of dense and sparse features, different optimizers, and tuning strategies for the CoNLL-2014 shared task. We notice erratic behavior when optimizing sparse feature weights with M2 and offer partial solutions. We find that a bare-bones phrase-based SMT setup with task-specific parameter-tuning outperforms all previously published results for the CoNLL-2014 test set by a large margin (46.37% M2 over previously 41.75%, by an SMT system with neural features) while being trained on the same, publicly available data. Our newly introduced dense and sparse features widen that gap, and we improve the state-of-the-art to 49.49% M2 .
1607.06153	In this paper, we present the first experiments using neural network models for the task of error detection in learner writing. We perform a systematic comparison of alternative compositional architectures and propose a framework for error detection based on bidirectional LSTMs. Experiments on the CoNLL-14 shared task dataset show the model is able to outperform other participants on detecting errors in learner writing. Finally, the model is integrated with a publicly deployed self-assessment system, leading to performance comparable to human annotators.
1707.05236	Shortage of available training data is holding back progress in the area of automated error detection. This paper investigates two alternative methods for artificially generating writing errors, in order to create additional resources. We propose treating error generation as a machine translation task, where grammatically correct text is translated to contain errors. In addition, we explore a system for extracting textual patterns from an annotated corpus, which can then be used to insert errors into grammatically correct sentences. Our experiments show that the inclusion of artificially generated errors significantly improves error detection accuracy on both FCE and CoNLL 2014 datasets.
E14-3013	This paper explores the generation of artificial errors for correcting grammatical mistakes made by learners of English as a second language. Artificial errors are injected into a set of error-free sentences in a probabilistic manner using statistics from a corpus. Unlike previous approaches, we use linguistic information to derive error generation probabilities and build corpora to correct several error types, including open-class errors. In addition, we also analyse the variables involved in the selection of candidate sentences. Experiments using the NUCLE corpus from the CoNLL 2013 shared task reveal that: 1) training on artificially created errors improves precision at the expense of recall and 2) different types of linguistic information are better suited for correcting different error types.
W14-1702	This paper describes our submission to the CoNLL 2014 shared task on grammatical error correction using a hybrid approach, which includes both a rule-based and an SMT system augmented by a large

	webbased language model. Furthermore, we demonstrate that correction type estimation can be used to remove unnecessary corrections, improving precision without harming recall. Our best hybrid system achieves state-of-the-art results, ranking first on the original test set and second on the test set with alternative annotations.
W07-1604	This paper presents ongoing work on the detection of preposition errors of non-native speakers of English. Since prepositions account for a substantial proportion of all grammatical errors by ESL (English as a Second Language) learners, developing an NLP application that can reliably detect these types of errors will provide an invaluable learning resource to ESL students. To address this problem, we use a maximum entropy classifier combined with rule-based filters to detect preposition errors in a corpus of student essays. Although our work is preliminary, we achieve a precision of 0.8 with a recall of 0.3.
I17-1005	In this study, we improve grammatical error detection by learning word embeddings that consider grammaticality and error patterns. Most existing algorithms for learning word embeddings usually model only the syntactic context of words so that classifiers treat erroneous and correct words as similar inputs. We address the problem of contextual information by considering learner errors. Specifically, we propose two models: one model that employs grammatical error patterns and another model that considers grammaticality of the target word. We determine grammaticality of n-gram sequence from the annotated error tags and extract grammatical error patterns for word embeddings from large-scale learner corpora. Experimental results show that a bidirectional long-short term memory model initialized by our word embeddings achieved the state-of-the-art accuracy by a large margin in an English grammatical error detection task on the First Certificate in English dataset.
han-lrec10-final	This paper presents research on building a model of grammatical error correction, for preposition errors in particular, in English text produced by language learners. Unlike most previous work which trains a statistical classifier exclusively on well-formed text written by native speakers, we train a classifier on a large-scale, error-tagged corpus of English essays written by EFL learners, relying on contextual and grammatical features surrounding preposition usage. First, we show that such a model can achieve high performance values: 93.3% precision and 14.8% recall for error detection and 81.7% precision and 13.2% recall for error detection and correction when tested on preposition replacement errors. Second, we show that this model outperforms models trained on well-edited text produced by native speakers of English. We discuss the implications of our approach in the area of language error modeling and the issues stemming from working with a noisy data set whose error annotations are not exhaustive.
P11-2089	Despite the rising interest in developing grammatical error detection systems for non-native speakers of English, progress in the field has been hampered by a lack of informative metrics and an inability to directly compare the performance of systems developed by different researchers. In this paper we address these problems by presenting two evaluation methodologies, both based on a novel use of crowdsourcing.

C08-1022	In this paper, we present an approach to the automatic identification and correction of preposition and determiner errors in nonnative (L2) English writing. We show that models of use for these parts of speech can be learned with an accuracy of 70.06% and 92.15% respectively on L1 text, and present first results in an error detection task for L2 writing.
D14-1102	Different approaches to high-quality grammatical error correction have been proposed recently, many of which have their own strengths and weaknesses. Most of these approaches are based on classification or statistical machine translation (SMT). In this paper, we propose to combine the output from a classification-based system and an SMT-based system to improve the correction quality. We adopt the system combination technique of Heafield and Lavie (2010). We achieve an F0.5 score of 39.39% on the test set of the CoNLL-2014 shared task, outperforming the best system in the shared task.
W11-2838	The aim of the Helping Our Own (HOO) Shared Task is to promote the development of automated tools and techniques that can assist authors in the writing task, with a specific focus on writing within the natural language processing community. This paper reports on the results of a pilot run of the shared task, in which six teams participated. We describe the nature of the task and the data used, report on the results achieved, and discuss some of the things we learned that will guide future versions of the task.
1804.05945	We combine two of the most popular approaches to automated Grammatical Error Correction (GEC): GEC based on Statistical Machine Translation (SMT) and GEC based on Neural Machine Translation (NMT). The hybrid system achieves new state-of-the-art results on the CoNLL-2014 and JFLEG benchmarks. This GEC system preserves the accuracy of SMT output and, at the same time, generates more fluent sentences as it typical for NMT. Our analysis shows that the created systems are closer to reaching human-level performance than any other GEC system reported so far.
W17-5037	We build a grammatical error correction (GEC) system primarily based on the state-of-the-art statistical machine translation (SMT) approach, using task-specific features and tuning, and further enhance it with the modeling power of neural network joint models. The SMT-based system is weak in generalizing beyond patterns seen during training and lacks granularity below the word level. To address this issue, we incorporate a character-level SMT component targeting the misspelled words that the original SMT-based system fails to correct. Our final system achieves 53.14% F0.5 score on the benchmark CoNLL-2014 test set, an improvement of 3.62% F0.5 over the best previous published score.
gleu_update_2016	The GLEU metric was proposed for evaluating grammatical error corrections using n-gram overlap with a set of reference sentences, as opposed to precision/recall of specific annotated errors (Napoles et al., 2015). This paper describes improvements made to the GLEU metric that address problems that arise when using an increasing number of reference sets. Unlike the originally presented metric, the modified metric does not require tuning.
1801.08831	We improve automatic correction of grammatical, orthographic, and collocation errors in text using a multilayer convolutional encoder-decoder neural network. The network is initialized with embeddings that make use of character Ngram information to better suit this task. When evaluated on common benchmark test data sets

	(CoNLL-2014 and JF- LEG), our model substantially outperforms all prior neural approaches on this task as well as strong statistical machine translation-based systems with neural and task-specific features trained on the same data. Our analysis shows the superiority of convolutional neural networks over recurrent neural networks such as long short-term memory (LSTM) networks in capturing the local context via attention, and thereby improving the coverage in correcting grammatical errors. By ensembling multiple models, and incorporating an N-gram language model and edit features via rescoring, our novel method becomes the first neural approach to outperform the current state-of-the-art statistical machine translation-based approach, both in terms of grammaticality and fluency.
D16-1195	An important aspect for the task of grammatical error correction (GEC) that has not yet been adequately explored is adaptation based on the native language (L1) of writers, despite the marked influences of L1 on second language (L2) writing. In this paper, we adapt a neural network joint model (NNJM) using L1-specific learner text and integrate it into a statistical machine translation (SMT) based GEC system. Specifically, we train an NNJM on general learner text (not L1-specific) and subsequently train on L1-specific data using a Kullback-Leibler divergence regularized objective function in order to preserve generalization of the model. We incorporate this adapted NNJM as a feature in an SMT-based English GEC system and show that adaptation achieves significant F0.5 score gains on English texts written by L1 Chinese, Russian, and Spanish writers.
1606.00189	Phrase-based statistical machine translation (SMT) systems have previously been used for the task of grammatical error correction (GEC) to achieve state-of-the-art accuracy. The superiority of SMT systems comes from their ability to learn text transformations from erroneous to corrected text, without explicitly modeling error types. However, phrase-based SMT systems suffer from limitations of discrete word representation, linear mapping, and lack of global context. In this paper, we address these limitations by using two different yet complementary neural network models, namely a neural network global lexicon model and a neural network joint model. These neural networks can generalize better by using continuous space representation of words and learn non-linear mappings. Moreover, they can leverage contextual information from the source sentence more effectively. By adding these two components, we achieve statistically significant improvement in accuracy for grammatical error correction over a state-of-the-art GEC system.
N16-1133	Research on grammatical error correction has received considerable attention. For dealing with all types of errors, grammatical error correction methods that employ statistical machine translation (SMT) have been proposed in recent years. An SMT system generates candidates with scores for all candidates and selects the sentence with the highest score as the correction result. However, the 1-best result of an SMT system is not always the best result. Thus, we propose a reranking approach for grammatical error correction. The reranking approach is used to re-score N-best results of the SMT and reorder the results. Our experiments show that our reranking system using parts of speech and syntactic features improves performance and achieves state-of-the-art quality, with an F0.5 score of 40.0.

W13-3607	This paper describes our use of phrasebased statistical machine translation (PB- SMT) for the automatic correction of errors in learner text in our submission to the CoNLL 2013 Shared Task on Grammatical Error Correction. Since the limited training data provided for the task was insufficient for training an effective SMT system, we also explored alternative ways of generating pairs of incorrect and correct sentences automatically from other existing learner corpora. Our approach does not yield particularly high performance but reveals many problems that require careful attention when building SMT systems for error correction.
N16-1042	This paper presents the first study using neural machine translation (NMT) for grammatical error correction (GEC). We propose a twostep approach to handle the rare word problem in NMT, which has been proved to be useful and effective for the GEC task. Our best NMT-based system trained on the CLC outperforms our SMT-based system when testing on the publicly available FCE test set. The same system achieves an F0.5 score of 39.90% on the CoNLL-2014 shared task test set, outperforming the state-of-the-art and demonstrating that the NMT-based GEC system generalises effectively.
W16-0530	We develop a supervised ranking model to rerank candidates generated from an SMT-based grammatical error correction (GEC) system. A range of novel features with respect to GEC are investigated and implemented in our reranker. We train a rank preference SVM model and demonstrate that this outperforms both Minimum Bayes-Risk and Multi-Engine Machine Translation based re-ranking for the GEC task. Our best system yields a significant improvement in I-measure when testing on the publicly available FCE test set (from 2.87% to 9.78%). It also achieves an F0.5 score of 38.08% on the CoNLL-2014 shared task test set, which is higher than the best original result. The oracle score (upper bound) for the re-ranker achieves over 40% I-measure performance, demonstrating that there is considerable room for improvement in the re-ranking component developed here, such as incorporating features able to capture long-distance dependencies.
1603.097 27	Natural language correction has the potential to help language learners improve their writing skills. While approaches with separate classifiers for different error types have high precision, they do not flexibly handle errors such as redundancy or non-idiomatic phrasing. On the other hand, word and phrase-based machine translation methods are not designed to cope with orthographic errors, and have recently been outpaced by neural models. Motivated by these issues, we present a neural network-based approach to language correction. The core component of our method is an encoder-decoder recurrent neural network with an attention mechanism. By operating at the character level, the network avoids the problem of out-of-vocabulary words. We illustrate the flexibility of our approach on dataset of noisy, user-generated text collected from an English learner forum. When combined with a language model, our method achieves a state-of-the-art F0:5-score on the CoNLL 2014 Shared Task. We further illustrate that training the network on additional data with synthesized errors can improve performance
P18-1097	Most of the neural sequence-to-sequence (seq2seq) models for grammatical error correction (GEC) have two limitations: (1) a seq2seq model may not be well generalized with only limited error-

	corrected data; (2) a seq2seq model may fail to completely correct a sentence with multiple errors through normal seq2seq inference. We attempt to address these limitations by proposing a fluency boost learning and inference mechanism. Fluency boosting learning generates fluency-boost sentence pairs during training, enabling the error correction model to learn how to improve a sentence's fluency from more instances, while fluency boosting inference allows the model to correct a sentence incrementally through multi-round seq2seq inference until the sentence's fluency stops increasing. Experiments show our approaches improve the performance of seq2seq models for GEC, achieving state-of-the-art results on both CoNLL2014 and JFLEG benchmark datasets.
1807.01270	Neural sequence-to-sequence (seq2seq) approaches have proven to be successful in grammatical error correction (GEC). Based on the seq2seq framework, we propose a novel fluency boost learning and inference mechanism. Fluency boosting learning generates diverse error-corrected sentence pairs during training, enabling the error correction model to learn how to improve a sentence's fluency from more instances, while fluency boosting inference allows the model to correct a sentence incrementally with multiple inference steps. Combining fluency boost learning and inference with convolutional seq2seq models, our approach achieves the state-of-the-art performance: 75.72 (F0:5) on CoNLL-2014 10 annotation dataset and 62.42 (GLEU) on JFLEG test set respectively, becoming the first GEC system that reaches human-level performance (72.58 for CoNLL and 62.37 for JFLEG) on both of the benchmarks.
1804.11225	Metric validation in Grammatical Error Correction (GEC) is currently done by observing the correlation between human and metric-induced rankings. However, such correlation studies are costly, methodologically troublesome, and suffer from low inter-rater agreement. We propose MAEGE, an automatic methodology for GEC metric validation, that overcomes many of the difficulties with existing practices. Experiments with MAEGE shed a new light on metric quality, showing for example that the standard M2 metric fares poorly on corpus-level ranking. Moreover, we use MAEGE to perform a detailed analysis of metric behavior, showing that correcting some types of errors is consistently penalized by existing metrics.
1606.00210	Grammatical error correction (GEC) is the task of detecting and correcting grammatical errors in texts written by second language learners. The statistical machine translation (SMT) approach to GEC, in which sentences written by second language learners are translated to grammatically correct sentences, has achieved state-of-the-art accuracy. However, the SMT approach is unable to utilize global context. In this paper, we propose a novel approach to improve the accuracy of GEC, by exploiting the n-best hypotheses generated by an SMT approach. Specifically, we build a classifier to score the edits in the n-best hypotheses. The classifier can be used to select appropriate edits or re-rank the n-best hypotheses. We apply these methods to a state-of-the-art GEC system that uses the SMT approach. Our experiments show that our methods achieve statistically significant improvements in accuracy over the best published results on a benchmark test dataset on GEC.

6.Results

<table>
<tr><td>Result</td><td>Paper ID</td></tr>
<tr><td>CoNLL 2014
<table>
<tr><th>Rank</th><th>TeamID</th><th>P</th><th>R</th><th>M2</th></tr>
<tr><td>0.5</td><td></td><td></td><td></td><td></td></tr>
<tr><td>1</td><td>CAMB</td><td>39.71</td><td>30.10</td><td>37.33</td></tr>
<tr><td>2</td><td>CUUI</td><td>41.78</td><td>24.88</td><td>36.79</td></tr>
<tr><td>3</td><td>AMU</td><td>41.62</td><td>21.40</td><td>35.01</td></tr>
<tr><td>4</td><td>POST</td><td>34.51</td><td>21.73</td><td>30.88</td></tr>
<tr><td>5</td><td>NTHU</td><td>35.08</td><td>18.85</td><td>29.92</td></tr>
<tr><td>6</td><td>RAC</td><td>33.14</td><td>14.99</td><td>26.68</td></tr>
<tr><td>7</td><td>UMC</td><td>31.27</td><td>14.46</td><td>25.37</td></tr>
<tr><td>8</td><td>PKU</td><td>32.21</td><td>13.65</td><td>25.32</td></tr>
<tr><td>9</td><td>NARA</td><td>21.57</td><td>29.38</td><td>22.78</td></tr>
<tr><td>10</td><td>SJTU</td><td>30.11</td><td>5.10</td><td>15.19</td></tr>
<tr><td>11</td><td>UFC</td><td>70.00</td><td>1.72</td><td>7.84</td></tr>
<tr><td>12</td><td>IPN</td><td>11.28</td><td>2.85</td><td>7.09</td></tr>
<tr><td>13</td><td>IITB</td><td>30.77</td><td>1.39</td><td>5.90</td></tr>
</table>
Table 10: Shared Task results for submission without alternative answers. AMU is our result.</td><td>[W14-1703]</td></tr>
</table>

7.Bibliography

[N10-1018]	Alla Rozovskaya and Dan Roth. 2010. Training paradigms for correcting errors in grammar and usage. In NAACL.
[P11-1093]	Alla Rozovskaya and Dan Roth. 2011. Algorithm selection and model adaptation for esl correction tasks. In ACL.
[P16-1208]	Alla Rozovskaya and Dan Roth. 2016. Grammatical error correction: Machine translation and classifiers. In ACL.
[W14-1704]	Alla Rozovskaya, Kai-Wei Chang, Mark Sammons, Dan Roth, and Nizar Habash. 2014. The Illinois-columbia system in the conll-2014 shared task. In CoNLL (Shared Task).
[W12-2032]	Alla Rozovskaya, Mark Sammons, and Roth Dan. 2012. The ui system in the hoo 2012 shared task on error correction. In Workshop on Building Educational Applications Using NLP.
[1707.09067]	Allen Schmaltz, Yoon Kim, Alexander Rush, and Stuart Shieber. 2017. Adapting sequence models for sentence correction. In EMNLP.
[I13-1122]	Bibek Behera and Pushpak Bhattacharyya. 2013. Automated grammar correction using hierarchical phrase-based statistical machine translation. In IJC-NLP.
[P06-1032]	Chris Brockett, William B Dolan, and Michael Gamon. 2006. Correcting ESL errors using phrasal SMT techniques. In COLING/ACL.
[P17-1074]	Christopher Bryant, Mariano Felice, and Ted Briscoe. 2017. Automatic annotation and evaluation of error types for grammatical error correction. In ACL.
[AutomatedGram]	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.

[1610.02124]	Courtney Napoles, Keisuke Sakaguchi, and Joel Tetreault. 2016. There's no comparison: Reference-less evaluation metrics in grammatical error correction. In EMNLP.
[1702.04066]	Courtney Napoles, Keisuke Sakaguchi, and Joel Tetreault. 2017. Jfleg: A fluency corpus and benchmark for grammatical error correction. arXiv preprint arXiv: 1702.04066.
[P15-2097]	Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel Tetreault. 2015. Ground truth for grammatical error correction metrics. In ACUIJCNLP.
[D11-1010]	Daniel Dahlmeier and Hwee Tou Ng. 2011. Correcting semantic collocation errors with 11-induced paraphrases. In EMNLP.
[D12-1052]	Daniel Dahlmeier and Hwee Tou Ng. 2012a. A beam-search decoder for grammatical error correction. In EMNLP/CoNLL.
[N12-1067]	Daniel Dahlmeier and Hwee Tou Ng. 2012b. Better evaluation for grammatical error correction. In NAACL.
[D17-1297]	Helen Yannakoudakis, Marek Rei, Oistein E Andersen, and Zheng Yuan. 2017. Neural sequence-labelling models for grammatical error correction. In EMNLP.
[I17-2058]	Hiroki Asano, Tomoya Mizumoto, and Kentaro Inui. 2017. Reference-based metrics can be replaced with reference-less metrics in evaluating grammatical error correction systems. In IJCNLP.
[W14-1701]	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 CoNLL (Shared Task).
[W13-3604]	Ippei Yoshimoto, Tomoya Kose, Kensuke Mitsuzawa, Keisuke Sakaguchi, Tomoya Mizumoto, Yuta Hayashibe, Mamoru Komachi, and Yuji Matsumoto. 2013. Naist at 2013 conll grammatical error correction shared task. In CoNLL (Shared Task).
[Foster-BEA4]	Jennifer Foster and Oistein E Andersen. 2009. Generrate: generating errors for use in grammatical error detection. In Workshop on innovative use of nlp for building educational applications.
[1707.02026]	Jianshu Ji, Qinlong Wang, Kristina Toutanova, Yongen Gong, Steven Truong, and Jianfeng Gao. 2017. A nested attention neural hybrid model for grammatical error correction. In ACL.
[P10-2065]	Joel Tetreault, Jennifer Foster, and Martin Chodorow. 2010. Using parse features for preposition selection and error detection. In ACL.
[W10-1006]	Joel Tetreault, Elena Filatova, and Martin Chodorow. 2010. Rethinking grammatical error annotation and evaluation with the amazon mechanical turk. In Workshop on Innovative Use of NLP for Building Educational Applications.
[Q16-1013]	Keisuke Sakaguchi, Courtney Napoles, Matt Post, and Joel Tetreault. 2016. Reassessing the goals of grammatical error correction: Fluency instead of grammaticality. Transactions of the Association of Computational Linguistics, 4(1):169-182.
[1707.00299]	Keisuke Sakaguchi, Matt Post, and Benjamin Van Durme. 2017. Grammatical error correction with neural reinforcement learning. In IJCNLP.
[W14-1703]	Marcin Junczys-Dowmunt and Roman Grundkiewicz. 2014. The amu system in the conll-2014 shared task: Grammatical error correction by data-intensive and feature-rich statistical machine translation. In CoNLL (Shared Task).
[1605.06353]	Marcin Junczys-Dowmunt and Roman Grundkiewicz. 2016. Phrase-based machine translation is state-of-the-art for automatic grammatical error correction. arXiv preprint arXiv: 1605.06353.
[1607.06153]	Marek Rei and Helen Yannakoudakis. 2016. Compositional sequence labeling models for error detection in learner writing. In ACL.

[1707.05236]	Marek Rei, Mariano Felice, Zheng Yuan, and Ted Briscoe. 2017. Artificial error generation with machine translation and syntactic patterns. arXiv preprint arXiv:1707.05236.
[E14-3013]	Mariano Felice and Zheng Yuan. 2014. Generating artificial errors for grammatical error correction. In Student Research Workshop at EACL.
[W14-1702]	Mariano Felice, Zheng Yuan, Oistein E Andersen, Helen Yannakoudakis, and Ekaterina Kochmar. 2014. Grammatical error correction using hybrid systems and type filtering. In CoNLL (Shared Task).
[W07-1604]	Martin Chodorow, Joel Tetreault, and Na-Rae Han. 2007. Detection of grammatical errors involving prepositions. In ACL-SIGSEM workshop on prepositions.
[I17-1005]	Masahiro Kaneko, Yuya Sakaizawa, and Mamoru Komachi. 2017. Grammatical error detection using error-and grammaticality-specific word embeddings. In IJCNLP.
[han-lrec10-final]	Na-Rae Han, Joel Tetreault, Soo-Hwa Lee, and Jin-Young Ha. 2010. Using an error-annotated learner corpus to develop an esl/efl error correction system. In LREC.
[P11-2089]	Nitin Madnani, Joel Tetreault, Martin Chodorow, and Alla Rozovskaya. 2011. They can help: Using crowdsourcing to improve the evaluation of grammatical error detection systems. In ACL.
[C08-1022]	Rachele De Felice and Stephen G Pulman. 2008. A classifier-based approach to preposition and determiner error correction in 12 english. In COLING.
[D14-1102]	Raymond Hendy Susanto, Peter Phandi, and Hwee Tou Ng. 2014. System combination for grammatical error correction. In EMNLP.
[W11-2838]	Robert Dale and Adam Kilgarriff. 2011. Helping our own: The hoo 2011 pilot shared task. In European Workshop on Natural Language Generation.
[1804.05945]	Roman Grundkiewicz and Marcin Junczys-Dowmunt. 2018. Near human-level performance in grammatical error correction with hybrid machine translation. arXiv preprint arXiv: 1804.05945.
[1804.05940]	Marcin Junczys-Dowmunt, Roman Grundkiewicz, Shubha Guha and Kenneth Heafield. Approaching Neural Grammatical Error Correction as a Low-Resource Machine Translation Task. 2018. arXiv preprint arXiv: 1804.05940.
[W17-5037]	Shamil Chollampatt and Hwee Tou Ng. 2017. Connecting the dots: Towards human-level grammatical error correction. In Workshop on Innovative Use of NLP for Building Educational Applications.
[gleu_update_2016]	Courtney Napoles, Keisuke Sakaguchi, Matt Post and Joel Tetreault. 2016. GLEU Without Tuning. Johns Hopkins University
[1801.08831]	Shamil Chollampatt and Hwee Tou Ng. 2018. A multi-layer convolutional encoder-decoder neural network for grammatical error correction. arXiv preprint arXiv: 1801.08831.
[D16-1195]	Shamil Chollampatt, Duc Tam Hoang, and Hwee Tou Ng. 2016a. Adapting grammatical error correction based on the native language of writers with neural network joint models. In EMNLP.
[1606.00189]	Shamil Chollampatt, Kaveh Taghipour, and Hwee Tou Ng. 2016b. Neural network translation models for grammatical error correction. arXiv preprint arXiv: 1606.00189.
[N16-1133]	Tomoya Mizumoto and Yuji Matsumoto. 2016. Discriminative reranking for grammatical error correction with statistical machine translation. In NAACL.
[W13-3607]	Zheng Yuan and Mariano Felice. 2013. Constrained grammatical error correction using statistical machine translation. In CoNLL (Shared Task).
[N16-1042]	Zheng Yuan and Ted Briscoe. 2016. Grammatical error correction using neural machine translation. In NAACL.
[W16-0530]	Zheng Yuan, Ted Briscoe, Mariano Felice, Zheng Yuan, Ted Briscoe, and Mariano Felice. 2016. Candidate re-ranking for

	smt-based grammatical error correction. In Workshop on Innovative Use of NLP for Building Educational Applications.
[1603.09727]	Ziang Xie, Anand Avati, Naveen Arivazhagan, Dan Jurafsky, and Andrew Y Ng. 2016. Neural language correction with character-based attention. arXiv preprint arXiv:1603.09727.
[P18-1097]	Tao Ge, Furu Wei and Ming Zhou. 2018. Fluency Boost Learning and Inference for Neural Grammatical Error Correction. ACL.
[1807.01270]	Tao Ge, Furu Wei and Ming Zhou. 2018. Reaching Human-level Performance in Automatic Grammatical Error Correction: An Empirical Study. arXiv preprint arXiv:1807.01270.
[1804.11225]	Leshem Choshen and Omri Abend. 2018. Automatic Metric Validation for Grammatical Error Correction. arXiv preprint arXiv: 1804.11225
[1606.00210]	Duc Tam Hoang, Shamil Chollampatt, and Hwee Tou Ng. 2016. Exploiting n-best hypotheses to improve an smt approach to grammatical error correction. In IJCAI.