

私のブックマーク

自然言語処理による文法誤り訂正^{†1}

水本 智也(理化学研究所革新知能総合研究センター)

1. はじめに

依頼を受けたときのテーマは「言語学習・教育支援」ということなのだが、このあたりのテーマについては甲南大学永田 亮氏の著書「語学学習支援のための言語処理 [1]」や Leacock らの「Automated Grammatical Error Detection for Language Learners[2]」にも詳しく書かれているため、広く知りたい方はそちらの本を読んでもらいたい。今回は特に「自然言語処理による英語の文法誤り訂正」に絞って最近の研究の動向やデータ・ツールを紹介し、すぐにでも文法誤り訂正を試してもらえる形で紹介したい。

あらためて説明する必要はないかもしれないが、念のため最初に文法誤り訂正のタスクについて説明しておく。一般的な文法誤り訂正では、非母語話者が書く文を入力として受け取り、その文に含まれる誤りを訂正した文を出力する。簡単な例を図 1 に示す。この例だと、本来は“to”とすべきところを“in”と誤って書いており、このような誤りを訂正するのが文法誤り訂正である。



図 1 文法誤り訂正の例

2. 文法誤り訂正の最近の動向

文法誤り訂正の研究自体は誤り検出(訂正するのではなく誤った箇所を検出する)を含め古くから行われてはいるが、ここでは主に 2011 年以降の研究動向について概観する。

2.1 文法誤り訂正のコンペティション (Shared Task)

文法誤り訂正の研究が盛んに行われるようになってきた理由の一つとして外すことができないのが、Shared Task と呼ばれるコンペティションの存在である。2011 年には Helping Our Own (HOO-2011 Shared Task) [3] と呼ばれるコンペティションが開催され、2012 年にも HOO-2012 Shared Task [4] が開催された。また、2012 年には Error Detection and Correction Workshop 2012 (EDCW 2012) [5] というコンペティションが日本で開催され、このあたりから文法誤り訂正の研究が盛んになってきていることがうかがえる。

2013 年、2014 年には、CoNLL と呼ばれる国際会議の Shared Task として文法誤り訂正が採用され (CoNLL-2013 Shared Task [6], CoNLL-2014 Shared Task [7])、ここで用意された共通の評価用データセットによって文法誤り訂正の研究が発展したといっても過言ではない、と筆者は思う。

以下にあげるそれぞれの Overview paper を読めば使用したデータ、評価方法、参加チームのシステムの性能などを知ることができるので、興味がある人は読んでみてほしい。

- Helping Our Own : The HOO 2011 Pilot Shared Task [8]
- HOO 2012 : A Report on the Preposition and Determiner Error Correction Shared Task [9]
- The CoNLL-2013 Shared Task on Grammatical Error Correction [10]
- The CoNLL-2014 Shared Task on Grammatical Error Correction [11]

2.2 文法誤り訂正の評価手法

文法誤り訂正研究の発展のうえで忘れてはいけないのが、文法誤り訂正の評価手法である。これらの評価手法があ

^{†1} http://www.ai-gakkai.or.jp/my-bookmark_vol33-no6

るために定量的評価が可能となりさまざまな手法の研究開発がスムーズに行われるようになっていく。

2012年以前の文法誤り訂正の評価は、1単語単位で適合率（訂正した単語が正しく訂正できているか）、再現率（訂正すべき箇所がどの程度正しく訂正できたか）、F値を計算していた。

2012年に Dahlmeier らによる“Better evaluation for grammatical error correction [12]”で Max Match (M2) という新たな評価手法が提案された。この評価手法はこれまで1単語単位で訂正が正しいかなどを見ていたものを、なるべく長い単位で一致を見ようとしたものであるが、基本的には従来どおり再現率、適合率を計算するが、適合率を重視した F 0.5 値で評価することが一般的になっている。CoNLL 2013 Shared Task と CoNLL 2014 Shared Task での評価指標として使われており、その後の論文ではこの評価手法が使われていることが一般的となっている。その後 2015 年に Felice らの“Towards a standard evaluation method for grammatical error detection and correction [13]”で I-measure という手法が提案された。この評価手法のモチベーションは、訂正した結果が元の文より悪くなっている場合にそれを反映したいということがあり、他の評価手法では 0 ~ 1 の値を取ることが多いが、I-measure では -1 ~ 1 の値を取るようになっていく。同じく 2015 年に Napoles らの“Ground truth for grammatical error correction metrics [14]”で GLEU という評価手法が提案された。名前から想像がつく方もいると思うが、これは機械翻訳の評価手法として使われる BLEU に着想を得てつくられている。機械翻訳では原文を見ることはないが、文法誤り訂正では原文、システム、正解の三つを比べる必要があり、そのあたりで BLEU から改良が加えられている。

この三つの評価手法について改めて評価を行った Chollampatt らの論文“A reassessment of reference-based grammatical error correction metrics [15]”が 2018 年に発表されているため、こちらも読むと面白い知見が得られると思う。

ここまでは正解データを使った文法誤り訂正の評価手法について紹介したが、正解データを使わずにシステムを評価しようという手法も提案されている。2016 年に Napoles らが“There’s no comparison: Reference-less evaluation metrics in grammatical error correction [16]”で初めて正解データを使わない評価手法を提案した。この論文では文法誤り検出（訂正ではなく誤っている箇所を検出する）システムで検出された誤りの数を使って文法誤り訂正のシステムを評価することを行っている。2017 年に Asano らが“Reference-based metrics can be replaced with reference-less metrics in evaluating grammatical error correction Systems [17]”で正解データを使わない新しい評価手法を提案した。この論文では、文法性・流暢性・意味保存性の三つの観点に注目した評価手法を提案しており、正解データを使わない評価手法が正解データを使った評価手法よりも人による評価に近い評価ができることを明らかにした。

評価手法自体ではないが、関連する研究として重要なものをいくつかピックアップして紹介する。CoNLL 2014 Shared Task の参加チームのシステムの出力 [18] を人手で評価した論文として 2015 年の Grundkiewicz らの“Human evaluation of grammatical error correction systems [19]”がある。その後の評価手法の性能を測る場合にこの論文で出されている人手の評価を使うことがあり、文法誤り訂正の評価に関する研究として押さえておきたい。また、文法誤り訂正においても流暢性が大切であることを示した Sakaguchi らの 2016 年の論文“Reassessing the goals of grammatical error correction: fluency instead of grammaticality [20]”も重要である。流暢性を重視して訂正されているデータもつくられているため、この論文も押さえておきたい。

2.3 文法誤り訂正の手法

文法誤り訂正の手法には、大きく分けてルールベース、分類器ベース、機械翻訳ベースの三つがある。

ルールベースの手法では、主に冠詞・名詞の単数複数・主語と動詞の一致・時制などの誤りを訂正するものが提案されている。2011 年以降はルールベースの手法はあまり行われてはいないが、主語と動詞の一致誤りを対象としたものとしては永田による 2013 年の論文“構文解析を必要としない主語動詞一致誤り検出手法 [21]”がある。また、前置詞を対象としたものとして Nagata らによる 2014 年の論文“Correcting preposition errors in learner english using error case frames and feedback messages [22]”があり、この手法は誤りを訂正するだけでなくフィードバックメッセージを出すことも可能で面白い。

2013 年までは分類器を使った文法誤り訂正が主流であり、訂正する誤りのタイプも冠詞や前置詞など一部の誤りタイプに限定されていた。一部のみを紹介すると、冠詞や前置詞を対象とした De Felice らの 2008 年の論文“A classifier-based approach to preposition and determiner error correction in L2 English [23]”，前置詞を対象としたものとして Dahlmeier らの 2011 年の論文“Grammatical error correction with alternating structure optimization [24]”や Rozovskaya らの 2011 年の論文“Algorithm selection and model adaptation for ESL correction tasks [25]”，動詞の時制を対象とした Tajiri らの 2012 年の論文“Tense and aspect error correction for ESL learners using global context [26]”などがある。動詞の語彙選択にフォーカスした Sawai らの 2013 年の論文“A learner

corpus-based approach to verb suggestion for ESL [27]”のようなものもある。

一番最初の機械翻訳ベースの文法誤り訂正の論文は、2006 年の Brockett らによる “Correcting ESL errors using phrasal SMT techniques [28]” である。この論文が出た当初は大規模な文法誤り訂正のデータがなかったため、擬似的に文法誤りの入ったデータを作成してモデルを学習した。大規模なデータがなかったためか、しばらくは機械翻訳ベースの文法誤り訂正の論文は出ていないが、2011 年に Mizumoto らが大規模な文法誤り訂正のデータを使った日本語の文法誤り訂正の論文 “Mining revision log of language learning SNS for automated Japanese error correction of second language learners [29]” を発表し、2012 年に英語を対象とした “The effect of learner corpus size in grammatical error correction of ESL writings [30]” も発表し、それ以降彼らのデータを使った機械翻訳ベースの手法が多く発表されている。

機械翻訳に詳しい人は知っていると思うが、機械翻訳には統計的機械翻訳と呼ばれる手法と、ニューラル機械翻訳と呼ばれる手法があり、文法誤り訂正においてもこの二つが用いられている。2016 年までは統計的機械翻訳が主流であり、それ以降はニューラル機械翻訳を使った文法誤り訂正が主流になってきている。ここではその一部を紹介する。

まず、統計的機械翻訳ベースの論文を二つ紹介する。

- Discriminative reranking for grammatical error correction with statistical machine translation [31]

Mizumoto らによる 2016 年の論文。統計的機械翻訳の訂正結果を複数出力し、それを別のシステムを使って並べ直す (リランキング) 手法を提案した論文である。言語モデルによるリランキングはこれまでもあったが、品詞などの情報も使ってリランキングした最初の論文である。また、統計的機械翻訳が出す複数の出力で一番良い訂正を選んだ場合のスコア (オラクルスコア) を調べている。

- Phrase-based machine translation is state-of-the-art for automatic grammatical error correction [32]

Junczys-Dowmunt らの 2016 年の論文。それまでの文法誤り訂正のシステムの性能 (F0.5 値) が 40 程度だったのに対し、この論文では F0.5 値で 52.21 を出し、その当時ダントツの性能を出した論文である。後述するがこちらはツール、モデルが公開されており、実際に動かすこともできる。

ニューラル機械翻訳を使った研究で押さえておきたいものとして五つ紹介する。

- Grammatical error correction using neural machine translation [33]

Yuan らによる 2016 年の論文。文法誤り訂正において初めてニューラル機械翻訳の手法を使った論文である。性能自体はそこまで高くはないが、初めてニューラル機械翻訳を文法誤り訂正に適応した論文として覚えておきたい。

- Grammatical error correction with neural reinforcement learning [34]

Sakaguchi らによる 2017 年の論文。ニューラル機械翻訳ベースの文法誤り訂正で初めて強化学習を使った論文である。この論文は、ここであげた他の論文と異なるデータセットでのみ評価しているため性能については割愛する (後述する JFLEG というデータで評価している)。

- A multilayer convolutional encoder-decoder neural network for grammatical error correction [35]

Chollampatt らによる 2018 年の論文。RNN 系のエンコーダ・デコーダを使った手法が多い中で CNN を使った手法を提案した。F0.5 値も 54.79 とその当時の最高性能であった。コードが公開されているため実際に動かすことが可能である。

- Approaching neural grammatical error correction as a low-resource machine translation task [36]

Junczys-Dowmunt らによる 2018 年の論文。RNN 系のエンコーダ・デコーダを使った場合に再現率が低くなる傾向にあったが、この論文ではいくつかの工夫を行うことで従来より高い再現率を達成している。F0.5 値でも 56.1 と高性能を出している。

- Fluency boost learning and inference for neural grammatical error correction [37]

Ge らによる 2018 年の論文。流暢性という観点に注目し、流暢な訂正ができるようにいくつか工夫することで訂正性能を向上させている。この論文で F0.5 値は 52.72 となっている (流暢性を考慮してない場合は F0.5 値 41.81)。その後、同じ著者らが arXiv で “Reaching human-level performance in automatic grammatical error correction: An empirical study [38]” という論文を公開しており、元の論文で RNN エンコーダ・デコーダを使っていたのを、こちらでは上の CNN エンコーダ・デコーダのモデルを使っており、F0.5 値で 62.42 まで達成しており、現状最も性能が高い論文である。

このほかにも分類器と統計的機械翻訳を一緒に使った手法や統計的機械翻訳とニューラル機械翻訳の両方を使った手法も提案されている。興味があればぜひ調べてほしい。

3. データセット

文法誤り訂正の分野でシステムやモデルの学習に使われているデータと評価に使われているデータを紹介する。

文法誤り訂正のシステムを学習したり評価するためのデータは、学習者の書いた英文とそれを訂正した文のペアになっている。訂正された箇所はどういう誤りか（例えば、“a” から “the” であれば冠詞誤りなど）が付けられたデータもあれば、何も情報は付いておらず素朴に機械翻訳の平行コーパスのように学習者の書いた英文とその訂正文のペアになっているだけのものもある。

ここでは、特に明確に区別はせずに文法誤り訂正の論文によく出てくるデータについて紹介する。また、学習用・評価用の区別は便宜上のものであり、論文によってはここであげた評価用のものを学習に使って別のコーパスで評価しているものもある。

3.1 学習用のデータ

機械学習や機械翻訳ベースの文法誤り訂正を行うためには大規模なデータが必要となる。下記に三つのコーパスをあげる。

- NAIST Lang-8 Learner Corpora [39]

これは奈良先端科学技術大学院大学の松本研究室から配布されているデータである。語学学習のための SNS Lang-8 [40] の添削ログからクローリングしてつくられたデータである。約 200 万文対からなり、文法誤り訂正用のデータとしては規模の大きいものとなっている。学習者の母語に関しても、日本人の英文も多いが他の言語を母語とする英文も数多く含まれている。

- NUS Corpus of Learner English (NUCLE) [41]

このデータはシンガポール国立大学のチームが開発したデータで、データの規模自体はおよそ 6 万文対で、NAIST Lang-8 Learner Corpora には劣るが、訂正の質自体は良いとされている。また、後述する現在の評価データとしてスタンダードとなっているコーパスと性質が似ているため、これを使って学習することが多い (NAIST Lang-8 Learner Corpora と一緒に用いられることが多い)。

- The EF-Cambridge Open Language Database (EFCAMDAT) [42]

Cambridge University のチームがつくったコーパスであり、訂正されていない文のペアも合わせると約 600 万文対、何らかの訂正がされているものだけでも約 250 万文対あり、文法誤り訂正のコーパスとしては非常に大規模なコーパスである。このコーパスを文法誤り訂正に使っている論文は今年 2018 年に 1 本出たデモペーパー (Cool English: A Grammatical Error Correction System Based on Large Learner Corpora [43]) のみであり、そのデータの性質などはあまりわかってはいないが、このデータも大規模であるため、今後使われることも多くなると思われる。データの詳細に関しては Automatic Linguistic annotation of large scale L2 databases : The EF-Cambridge Open Language Database (EFCAMDAT) [44] を参照してほしい。

3.2 評価用データ

文法誤り訂正の評価に使われているデータとして最もよく使われるものは、CoNLL 2014 Shared Task のテストセットであるが、他にいくつかの論文で使われている二つのデータをあげる。

- CoNLL 2014 Shared Task Test Set [45] (ダウンロードが始まるので注意)

2014 年に Shared Task が開催されて以降、このデータを使って評価している論文がほとんどであり、後述する M2Scorer を用いて評価される。また関連したデータとして、CoNLL 2013 Shared Task Test Set [46] (ダウンロードが始まるので注意) があり、システムやモデルのパラメータを決めるために使われることが多い。どちらも約 1 300 文対からなる。このコーパスの評価には後述する評価ツール M2Scorer が使われることが一般的である。

- Cambridge Learner Corpus First Certificate in English (CLC-FCE) [47]

Cambridge University でつくられたデータであり、Cambridge Learner Corpus (CLC) という大規模なコーパスのサブセットとなっている (CLC 本体のほうは基本的には共同研究などでしか使用できない)。データ規模的には約 3 万文対であり、他の評価用コーパスよりは規模が大きくこれをシステムの学習に使っている論文もある。

- JHU FLuency-Extended GUG (JFLEG) [48]

2016 年に Johns Hopkins University, Grammarly のメンバによってつくられたデータである。このデータが他のデータと比べて変わっている点として、他のデータが単純に文法誤りのみが訂正されているのに対し、JFLEG では文法誤りを訂正しつつなるべく流暢な表現に直すということを行っている点がある。このコー

パスは公開時点で作者が開発セットとテストセットにデータを分割しており、それぞれ約 750 文対からなる。このコーパスを評価に使う場合は後述する評価ツールの GLEU を使って評価することが一般的になっている。

4. ツール

4.1 文法誤り訂正のツール

計算機環境さえあればすぐにでも使うことのできるツールを二つ紹介する。

一つは統計的機械翻訳の手法である **Phrase-based Statistical machine translation** を使ったもので、もう一つはニューラル機械翻訳を使ったツールである。便宜上、各システムが提案された論文のタイトルを示す。

- **Phrase-based Machine Translation is State-of-the-Art for Automatic Grammatical Error Correction [49]**
このシステムは 2016 年に発表された統計的機械翻訳を使った文法誤り訂正の手法を使ったものであり、M2 Score の F0.5 値で 52.21 で比較的高い性能を出すことができる。
- **A Multilayer Convolutional Encoder-Decoder Neural Network for Grammatical Error Correction [50]**
ニューラル機械翻訳を使ったシステムとして公開されているものとしては、現状最も性能が高いシステムであり、F0.5 値では 54.79 である。

ツールではないが、現在最も高い性能を出しているシステムの出力結果が以下で公開されている。現在の文法誤り訂正システムの性能がどの程度か見てみたい場合はぜひ見てほしい。

- **Reaching human-level performance in automatic grammatical error correction : An empirical study [51]**

4.2 文法誤り訂正の評価のためのツール

自分でつくったシステムを評価するためには、評価するためのツールが必要である。

文法誤り訂正の動向で紹介した正解データを使わない評価手法はツールとしては公開されていないため、正解データを使う三つのツールについてのみあげておく。

- **M2Scorer [52]**
現在文法誤り訂正の評価で最もよく使われている M2 で評価することができるツール。このツールを使うためには正解のデータを特殊な形に変換する必要があるが、後述する **ERRANT** に含まれているツールを使うことでこのツールを使うための形に変換することができる。
- **GLEU [53]**
JFLEG をテストデータとして使った場合に使用されることが多い。M2Scorer や I-measure と異なり、システムの出力とその正解が単純なペアの形になっていれば使うことができるため、簡単に使用することができる。
- **I-measure [54]**
このツールを使って評価するためには、xml 形式に変換する必要がある。このツールに M2Scorer 用の形式から xml に変換するスクリプトが含まれている。

4.3 その他の便利なツール

文法誤り訂正の研究や開発をするうえで知っている便利なツールを紹介する。

- **ERRor ANnotation Toolkit (ERRANT) [55]**
このツールではシステムの出力に対して誤りのタイプを自動で付与することができる。例えば、“a”を“the”に訂正しているような場合は“冠詞誤り”を付与する。このツールを使うことで、自分のシステムがどの誤りタイプに強い・弱いかを知ることが可能である。自動推定精度は約 96%とされている。詳しくは“Automatic annotation and evaluation of error types for grammatical error correction [56]”や“Automatic extraction of learner errors in ESL sentences using linguistically enhanced alignments [57]”を参照してほしい。また、このツールには M2Scorer で評価可能な形式に変換するスクリプトも含まれている。

- **EFCamDat-Preprocess [58]**
文法誤り訂正のデータである EFCAMDAT を文法誤り訂正に使えるように処理するためのツール。
- **Lang8-NAIST-extractor [59]**
NAIST Lang-8 Corpora 用の抽出ツール。誤り文と正解後のペアを抽出することができる。NAIST Lang-8 Corpora を使った論文では、それぞれ独自の前処理が行われているため、論文を再現したい場合は論文に書いてある方法を使って抽出する必要があることに注意したい。

5. 国際会議

文法誤り訂正の研究についての最新の研究動向を知りたい場合は、自然言語処理のカンファレンスである

- ・ Annual Meeting of the Association for Computational Linguistics (ACL) [60]
- ・ Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) [61]
- ・ Conference on Empirical Methods in Natural Language Processing (EMNLP) [62]
- ・ International Conference on Computational Linguistics (COLING) [63]
- ・ International Joint Conference on Natural Language Processing (IJCNLP) [64]

あたりを調べるとよいが、年や会議によってばらつきがあり、多い場合もあれば0本の場合もあるため、文法誤り訂正についてサーベイする場合は頑張って探してほしい。

言語学習系の国際ワークショップとして以下のものがある。

- ・ Workshop on Innovative Use of NLP for Building Educational Applications (BEA) [65]
 - ・ Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA) [66]
- これらのワークショップには文法誤り訂正だけでなく自動採点や母語推定など自然言語処理を使った言語学習に関する研究が発表されている。

6. 実際に使えるサービス/デモ

ここまで文法誤り訂正の研究に関することを述べてきたが、最後に実際に動いているサービスやデモを紹介したい。実際に動いているものを触ってみて今の文法誤り訂正がどの程度のものか確かめてほしい。念のために言っておくと、実際のサービスで研究論文で発表されている手法が必ずしも使われているわけではないことに注意したい。

- ・ Grammarly [67]
- ・ Ginger [68]
- ・ LanguageTool [69]
- ・ GEC Cool English [70]

Grammarlyは有料会員になることで文法誤りだけでなく、表現なども一部指摘してくれる。また、他のツールなど違い誤りに対して簡単な説明も表示してくれる。他のツールなども書きたいことはあるが、公の場なので詳細は語らないため、気になる人はそれぞれのシステムで何が異なるか比べてみてほしい。

7. おわりに

本ブックマークでは、すぐにでも文法誤り訂正を試してもらえらることを目的として、文法誤り訂正の最近の動向からデータセットやツールについて紹介した。個人的にはかなり頑張って書いたほうだが、説明が足りていない箇所も多々あったり、重要な文献など抜けているかもしれないが、そのあたりはご容赦願いたい。最後に、このブックマークを読んで文法誤り訂正に取り組んで人が増えてくれると大変うれしく思う。

- [1] <http://www.coronasha.co.jp/np/isbn/9784339027617/>
- [2] <https://www.morganclaypool.com/doi/abs/10.2200/S00562ED1V01Y201401HLT025>
- [3] <https://www.mq.edu.au/research/research-centres-groups-and-facilities/innovative-technologies/centres/centre-for-language-technology-clt/research/projects/hoo-helping-our-own/hoo-2011>
- [4] <https://www.mq.edu.au/research/research-centres-groups-and-facilities/innovative-technologies/centres/centre-for-language-technology-clt/research/projects/hoo-helping-our-own/hoo-2012-shared-task-on-preposition-and-determiner-error-correction>
- [5] <https://sites.google.com/site/edcw2012/>
- [6] <https://www.comp.nus.edu.sg/~nlp/con1113st.html>
- [7] <https://www.comp.nus.edu.sg/~nlp/con1114st.html>
- [8] https://www.mq.edu.au/__data/assets/pdf_file/0008/78938/HOO2011FinalReport.pdf

- [9] <http://anthology.aclweb.org/W/W12/W12-2006.pdf>
- [10] <https://www.comp.nus.edu.sg/~nlp/conl113st/CoNLLST01.pdf>
- [11] <https://www.comp.nus.edu.sg/~nlp/conl114st/CoNLLST01.pdf>
- [12] <http://www.aclweb.org/anthology/N12-1067>
- [13] <http://www.aclweb.org/anthology/N15-1060>
- [14] <http://www.aclweb.org/anthology/P/P15/P15-2097.pdf>
- [15] <http://aclweb.org/anthology/C18-1231>
- [16] <https://aclweb.org/anthology/D16-1228>
- [17] <http://www.aclweb.org/anthology/I17-2058>
- [18] https://www.comp.nus.edu.sg/~nlp/conl114st/official_submissions.tar.gz
- [19] <http://www.aclweb.org/anthology/D15-1052>
- [20] <http://aclWeb.org/anthology/Q16-1013>
- [21] http://search.ieice.org/bin/pdf.php?lang=J&year=2013&fname=j96-d_5_1346&abst=j
- [22] <http://aclweb.org/anthology/P14-1071>
- [23] <http://www.aclweb.org/anthology/C08-1022>
- [24] <http://www.aclweb.org/anthology/P11-1092>
- [25] <http://www.aclweb.org/anthology/P11-1093>
- [26] <http://www.aclweb.org/anthology/P12-2039>
- [27] <http://www.aclweb.org/anthology/P13-2124>
- [28] <http://www.aclweb.org/anthology/P06-1032>
- [29] <http://www.aclweb.org/anthology/I11-1017>
- [30] http://tomo-wb.jp/paper/COLING2012_Mizumoto.pdf
- [31] <http://www.aclweb.org/anthology/N16-1133>
- [32] <https://aclweb.org/anthology/D16-1161>
- [33] <http://www.aclweb.org/anthology/N16-1042>
- [34] <http://www.aclweb.org/anthology/I17-2062>
- [35] <https://www.aaii.org/ocs/index.php/AAAI/AAAI18/paper/view/17308/16137>
- [36] <http://aclweb.org/anthology/N18-1055>
- [37] <http://aclweb.org/anthology/P18-1097>
- [38] <https://arxiv.org/abs/1807.01270>
- [39] <https://sites.google.com/site/naistlang8corpora/>
- [40] <http://lang-8.com/>
- [41] <https://www.comp.nus.edu.sg/~nlp/conl114st.html>
- [42] https://corpus.mml.cam.ac.uk/efcamdat2/public_html/
- [43] <http://aclweb.org/anthology/C18-2018>
- [44] <http://people.ds.cam.ac.uk/ta259/SLRF2013Geertzenetal.pdf>
- [45] <https://www.comp.nus.edu.sg/~nlp/conl114st/conl114st-test-data.tar.gz>
- [46] <https://www.comp.nus.edu.sg/~nlp/conl113st/release2.3.1.tar.gz>
- [47] <https://ilexir.co.uk/datasets/index.html>
- [48] <https://github.com/keisks/jfleg>
- [49] <https://github.com/grammatical/baselines-emnlp2016>
- [50] <https://github.com/nusnlp/mlconvgec2018>
- [51] <https://github.com/getao/human-performance-gec>
- [52] <https://github.com/nusnlp/m2scorer>
- [53] <https://github.com/cnap/gec-ranking>
- [54] <https://github.com/mfelice/imeasure>
- [55] <https://github.com/chrisjbryant/errant>
- [56] <http://aclweb.org/anthology/P/P17/P17-1074.pdf>
- [57] <http://aclweb.org/anthology/C/C16/C16-1079.pdf>
- [58] <https://github.com/NTHU-NLPLAB/EFCamDat-Preprocess>

- [59] <https://github.com/tomo-wb/Lang8-NAIST-extractor>
- [60] <https://acl2018.org>
- [61] <http://naacl2018.org>
- [62] <http://emnlp2018.org>
- [63] <https://coling2018.org>
- [64] <http://ijcnlp2017.org/site/page.aspx?pid=901&sid=1133&lang=en>
- [65] <http://www.cs.rochester.edu/~tetreaul/naacl-bea13.html>
- [66] <https://sites.google.com/view/nlp2018>
- [67] <https://www.grammarly.com/>
- [68] <https://www.gingersoftware.com/>
- [69] <https://languagetool.org/>
- [70] <https://nlp-ultron.cs.nthu.edu.tw/gec/>