

文法誤り訂正への訂正重要度の導入

永田亮¹ 高村大也²

¹ 甲南大学 ² 産業技術総合研究所

nagata-nlp2022 @ hyogo-u.ac.jp., takamura.hiroya@aist.go.jp

概要

本稿では、文法誤り訂正における訂正の重要度を定量化することを試みる。具体的には、書き手の意図が正しく伝わるかどうかという観点で訂正の重要度を定量化する手法を提案する。提案手法を作例と学習者コーパスに適用し、有効性と改善すべき点を議論する。

1 はじめに

文法誤りには様々なものが存在するため、訂正の目的に応じて、文法誤り訂正の重要度を定量化することは有益である。本稿では、書き手の意図が正確に伝わるということを重視して、訂正の重要度を定量化することを試みる。具体的には、訂正前後で伝わる意味の差異が大きいほど重要な訂正とみなすような重要度を定量化する方法を提案する。このように定義された訂正の重要度は、語学学習に応用できる可能性がある。例えば、書き手の意図の伝達を大きく阻害する誤りを率先して訂正し、学習者に提示するような応用が考えられる。また、訂正の重要度で各誤りを重みづけて誤り訂正システムの性能を評価するという応用にもつながる。

本稿では、次のような非常にシンプルなアイデアに基づいて訂正の重要度を定量化する。深層学習ベースの言語モデルを用いると文中の単語をベクトル(分散表現)に変換できる。分散表現は、単語の文中での意味を表すとみなせる¹⁾。したがって、訂正前後の単語それぞれから分散表現を得て、その差異を測ることで意味のズレを定量化する。このズレが、上述の「意味の差異」に対応するとする。

このようなアイデアに基づいた提案手法の有効性を作例と学習者コーパスを用いて吟味する。例え

ば、この手法は、“He read $a \rightarrow \phi$ books.”という訂正²⁾は“He read $an \rightarrow a$ book.”よりも重要と判定する。ただし、3節で述べるように、このシンプルなアイデアそのままでは、いくつかの問題が生じるため、本稿では、その解決方法も議論する。

2 関連研究

誤り訂正の重要度に関連した知見に「自然な習得順序」という仮説がある。この仮説は、第二言語習得では、学習者に共通する習得順序があると主張する。逆に言えば、順序を入れ替えて新たな項目を習得できないと解釈できる。この仮説が正しいとすると、次に習得される項目に関する誤りを訂正することが重要となる。しかしながら、この仮説を支持／否定する両方の研究成果があり、どちらが正しいのかということは未解決問題である(これらの研究については、文献[1]が詳しい)。

文法誤り検出／訂正システムの評価を目的とした、訂正難易度の定量化[2]や誤り情報の信頼性推定[3]に関する研究もある。訂正難易度や誤り情報の信頼性はある種の重要度と解釈することもできるが、本研究で目指す書き手の意図が正確に伝わるということを考慮する訂正の重要度とは異なる。

3 訂正重要度の定量化手法

訂正前後の意味を分散表現として表すことが提案手法の基礎となる。以降では、訂正前後に対応する分散表現をベクトル e と c でそれぞれ表記する。この e と c が、訂正前後の文の意味を精度よく表現していれば、 e と c の差異に基づいて、1節で述べた訂正の重要度を定量化できる。二つのベクトルの差異は、ベクトル間の余弦類似度で定量化することが多く、本稿でも採択する。ただし、値が高いほど重要であるという尺度にするために、1から余弦類似度を引いた値を訂正の重要度とする。

2) 矢印前後で訂正前後の単語列を表している。また、 ϕ は空単語で、単語の削除(と追加)を表すのに使用する。

1) 厳密にいうと、言語モデルから得られる分散表現が意味を表しているかは定かでない。しかしながら、文や単語の分散表現は意味(のようなもの)を捉えているということが経験的に知られており、本稿でも近似的に分散表現が意味を表しているとみなす。

ここで問題となるのは、 e と c をどのように得るかということである。様々な方法が考えられるが、本稿では、次の二つの方法を検討する。基本的に、言語モデル（具体的には BERT）から得られる分散表現を用いる。

一つ目の方法では、単語の分散表現を e 、 c とする。具体的には、(1) 訂正前後の文をそれぞれ独立に BERT に入力し、最終層の出力を得る、(2) 出力として得られる分散表現のうち訂正前後の単語列に対応するものを、それぞれ e 、 c とする。ただし、複数の単語からなる単語列の場合は平均をとる。このようにして得られた分散表現間の余弦類似度は訂正前後の単語列が表す意味の差に対応すると解釈できる。

この方法は直感的ではあるが、余剰誤りと抜け誤りの場合に問題が生じる。例えば、“We know **about**→ ϕ the issue.” では単語 **about** が余剰であり、訂正文では削除されている。このように削除された単語列については、どのように分散表現を求めたらよいかは自明ではない。同様のことが、抜けの誤りの場合に訂正前の文で起こる。本稿では、余剰／抜け単語の左右の単語に対する分散表現の平均をとることとする。

二つ目の方法は、この問題をより直接的に解決する。BERT では、一文全体に対応する分散表現が得られる。すなわち、特殊トークン [CLS] に対応する分散表現を e と c とする。ただし、個々の訂正の重要度を定量化するために、一文内に複数の誤りがある場合でも、一度に訂正する誤りは一つとする。この方法では、訂正前後の単語列に直接着目するのではなく、訂正前後で文全体の意味がどのように変化したかに基いて訂正の重要度を定量化する。

以上の通り、提案手法では、訂正前後の情報から得られるベクトル e と c に基いて訂正の重要度を定量化しようとする。仮に、 e や c で単語や文の意味を表すことができるとして、訂正前の文が表す意味とは何かという問題が生じる。例えば、“He read an book.” というのは非標準であり、意味をなさないと言主張することもできる。逆に、訂正できるということは、訂正前の文でも訂正後の文と同じ意味が伝わると主張することもできる。提案手法では、次のような仮定と近似を暗に行っている。まず、誤りを含む文が表す仮想的な意味が存在し、空間内のある一点に対応づけられると仮定している。同様に、訂正された文の意味も同一空間内の点に対応付けられ

ると仮定している。（この二点間が十分に近い場合、人間は前者から後者を特定できるが）、提案手法では、この二点間の差異を近似的に訂正の重要度としていると解釈できる。

4 使用データ

提案手法の有効性を定性的に分析するために内省に基いて訂正の重要度が異なる例文を作成した。表 1 に結果を示す。表 1 では、訂正のカテゴリごとに作例を分けし、訂正の重要度が高いと思われる順に並べている。カテゴリと重要度の並べ替えの詳細は付録 A に示す通りである。

作例に加えて、二種類の学習者コーパスも用いた。一つ目は、CoNLL-2013 Shared Task [4] のテストデータとした。分析の容易さを考えて、限定詞、名詞の数、主語動詞の一致、前置詞、動詞の形態に関する 5 種類の誤りのみがアノテーションされたものを用いた。二つ目は、ICNALE [5] に、独自に誤り情報を付与した 1,894 文を用いた。誤り体系 [6] (20 種の誤り) を用いた。

5 訂正重要度の定性分析

5.1 作例を対象にした場合

表 1 の作例に対して、提案手法を適用して訂正の重要度を求めた。表の第 3, 4 カラムが文ベースの手法と単語ベースの手法の結果にそれぞれ対応する。括弧内の数は、内省の順位との差を表す。

文ベースの手法の結果は内省の結果に非常に類似する。異なるのは動詞に関する誤りのみである。内省では、**listened to** と **heard** とで表す意味が異なるため重要度が一番高いとしたが、提案手法ではカテゴリ内でもっとも重要度が低くなっている。訂正前後とも文法的には正しく、また、「聞く」という基本の意味は共通するため重要度が低くなった可能性がある。他のカテゴリについては、内省による重要度順と同一である。例えば、直感的には、**a**→ ϕ **books** と **a books**→**book** は同程度重要と感じられるが、推定された重要度も類似した値になっている（ただし、重要度の差異をどのように解釈すべきかは別の問題であり注意が必要である）。また、“We discussed **about**→ ϕ the issue.” は、重要度はカテゴリ内で最も低い。確かに、この誤りは意味が伝わるという観点からは致命的な誤りでないともいえる。もちろん、母語話者らしい英語という観点からは重要な訂正で

表1 内省に基づいて作成した例文と訂正重要度の推定結果。

カテゴリ	作例	文ベース重要度 (順位差)	単語ベース重要度 (順位差)
動詞	He <u>listened to</u> → <u>heard</u> the piano.	0.069 (↓2)	0.290 (↓2)
	He <u>play</u> → <u>played</u> the piano.	0.162 (↑1)	0.330 (↑1)
	He <u>play</u> → <u>plays</u> the piano.	0.121 (↑1)	0.298 (↑1)
限定詞/ 数	He read <u>a</u> → <u>φ</u> books.	0.030 (0)	0.586 (0)
	He read a <u>books</u> → <u>book</u> .	0.029 (0)	0.151 (0)
	He read <u>an</u> → <u>a</u> book.	0.003 (0)	0.066 (0)
前置詞	We know <u>about</u> → <u>φ</u> the issue.	0.013 (0)	0.377 (0)
	We agree <u>φ</u> → <u>to</u> the proposal.	0.011 (0)	0.293 (↓1)
	We discussed <u>about</u> → <u>φ</u> the issue.	0.009 (0)	0.355 (↑1)

あるが、ここで強調したいのは、このように訂正の重要度を定量化することで文法誤り訂正を新しい観点から議論できる可能性があるということである。

一方、単語ベースの手法では順位の入替わりが散見される。特に余剰の誤りの訂正が重要とみなされる傾向にある。例えば、“He read a→φ books.”が全訂正中で最も重要度が高い。これは、3節で述べたように余剰／抜け誤りでは、比較する単語に問題があることに起因する。この例では、aとread booksに対する分散表現を比較している。この比較は直感に合わず、改善の余地がある。この点において、文の分散表現に基いた手法のほうがより自然に訂正の重要度を定量化しているといえる。このことを考慮して、以降では文ベースの手法の結果を報告する。

5.2 学習者コーパスを対象にした場合

表2に、CoNLL-2013における重要度の上位／下位5件を示す³⁾。なお、学習者コーパスを対象にした場合は、一文中に複数の誤りが含まれる場合があり、表2も含めて、以降の結果では、一見適切でないように思える訂正もあることに注意されたい。例えば、表2の最終行は、全体として、a problem of the health of the→an unhealthy societyと訂正されている。

表2では、大きな傾向としては、冠詞と名詞の数に関する誤りが最も重要と認識されている。特に、定冠詞の有無は、一般への言及(2, 4行目)と特定物への言及(1, 5行目)という違いがあり、伝わる意味を大きく左右することが多い。このことを(少なくとも表2の範囲で)定量化結果は反映している。逆に、主語動詞の一致、前置詞の誤りは重要度

が低い。例えば、available for→toの重要度は最も低い部類であるが、forでも書き手の意図は十分に伝わる。また、冠詞の誤りであっても、a/anの使い分けについては重要度は低い。ただし、上の議論で重要度が高いと分析した訂正(例：名詞の単複)でも重要度が低いと認識されている例もある。これは後述するように、文長が影響している可能性がある。

表3に、ICNALEを対象にした場合の上位10件を示す(紙面の関係から、下位10件は付録Bに示す)。この結果も、表2と同じような傾向を見て取ることができる。加えて、文意において重要な役割を果たす要素(主語、目的語、接続詞)が上位に来ている。また、couldの有無から生じる差異を捉えているところも興味深い。

以上の結果を要約すると、提案手法は、誤りのために書き手の意図とは異なる意味が伝わる、または、複数の解釈が生じる(例：a books；a bookかbooksで曖昧)ケースを重要であるとみなす傾向にある。一方、a/anの使い分けや主語動詞の不一致など形式的な誤りは重要度を低く見積もる。どちらのケースがより重要かは目的によるが、意味が伝わるということを重視した場合は、提案手法には所望の重要度を定量化できる可能性がある。しかしながら、提案手法は非常にシンプルなアイデアに基づいているため、いくつかの改善すべき点がある。6節でこの点について考察する。

6 提案手法の改善すべき点

前節までに示した定量化の結果をみると、重要度が上位の文は文長が短くなる傾向にあることがわかる。これは、文が長くなるにつれて、一つの誤りが文全体の意味に及ぼす影響が小さくなるためと予想できる。すなわち、訂正前後の文長が重要度に影響

3) 表2では、重要度が0と表記されているものもあるが、厳密には0ではなくごく小さい値を丸めた結果である。

表2 CoNLL-2013 (5種類の誤りタイプ)における訂正重要度(上位/下位5件).

重要度	訂正文
0.140	In $\phi \rightarrow$ the supermarket monitor is needed because we have to track thieves.
0.090	The police seize the $\rightarrow \phi$ criminal thanks to CCTV.
0.088	In addition, tracking people system \rightarrow systems may be used by bad people in a wrong way.
0.087	The air cargo of the $\rightarrow \phi$ Valujet plane was on fire after the plane had taken off.
0.087	In $\phi \rightarrow$ the late nineteenth century, there was a severe air crash happening at Miami international airport.
0	The readers are just smiling when they flip the page because it never comes to their mind \rightarrow minds that one day it might come true.
0	..., employees will dislike and claim that their humans' right and privacy have \rightarrow has been violated.
0	... such powerful device shall not be made easily available for \rightarrow to the public or fall into the hands of 'ill-intentioned' individuals.
0	I agree to a great extent that RFID technology shall not be easily available for \rightarrow to the public and taken for granted.
0	Close surveillance is not only a problem of bad feelings, but also a \rightarrow an problem of the health of the society.

表3 ICNALEにおける訂正重要度(上位10件).

重要度	訂正文
0.251	$\phi \rightarrow$ A Restaurant is not correct place to smoking.
0.226	Restaurant is not $\phi \rightarrow$ the correct place to smoking.
0.182	... has worst result in besides non-smoker $\phi \rightarrow$, too.
0.179	... have many positive $\phi \rightarrow$ aspects for people ...
0.154	... cigarette has \rightarrow the worst result in ...
0.154	I could $\rightarrow \phi$ thank to my parents as I knew ...
0.154	When \rightarrow Once I smoke cigarette smoke, ...
0.150	Part time job \rightarrow jobs are for short time ...
0.148	$\phi \rightarrow$ You can Work your own schedule.
0.148	$\phi \rightarrow$ They Know their roles and responsibilities.

する可能性がある。同様に、(今回の結果には顕著には見られないが)、訂正前後の単語列の長さも影響を及ぼす可能性がある。

また別の問題として、一文中にいくつの誤りが存在するかということもある。提案手法では、一度に一つの誤りのみを訂正して重要度を算出しているが、このことは理論上、次の問題を引き起こす可能性がある。例えば、表2の7行目や最終行のように、一度に一つの誤りのみを訂正すると、不自然な文となる場合がある。このような不自然な文と原文を比較して得られる余弦類似度の差異は、伝わる意味の差異に適切に対応しない可能性が高い。また、一文中の誤りの数が増えると文全体としての意味の差異は大きくなると予想されるが、全体としての差異をどのように個々の誤りに還元するかということも問題になる。更に、このことは、誤りの単位をどのように設定するかという別の問題にもつながる。例えば、the books を a book に訂正した場合、誤り数2(冠詞と名詞の数の誤り)とすることも、全体と

して1つとすることもできる。この例では、直感的には全体として一つの誤りとして重要度を定量化すればよいように思われるが、別の例ではその限りではない(例: the book has を the books have に訂正した場合)。

以上をまとめると、提案手法には、少なくとも次の3つの改善すべき点があるといえる:

1. 訂正前後の文長から受ける影響の低減
2. 訂正前後の単語列の長さから受ける影響の低減
3. 複数の誤りの取り扱い

より実用的な訂正重要度の実現のためには、これらの要因について調整が必要である。例えば、文長や訂正前後の単語列の長さで正規化することなどが考えられる⁴⁾。また、複数の誤りの取り扱いについては、連続した誤りを一つにまとめ、訂正前後とも訂正箇所の左右の単語に対応する分散表現を使用することなどが考えられる。更に、推定された訂正の重要度をどのように定量評価するかということも今後の課題となる。

7 おわりに

本稿では、文法誤り訂正に、訂正の重要度という新たな観点を導入した。訂正前後の文から得られる分散表現に基いて訂正の重要度を定量化する手法を提案した。作例および実際の学習者コーパスに適用し、有効性と改善すべき点を議論した。定性分析の結果、定量化された訂正の重要度は直感に合う部分も見られたが、実用のためには、改善すべき点があることも確認された。

4) 試しに、訂正後の文長を対数にしたもので正規化したところ、正規化の影響が強すぎる傾向がみられた。

謝辞

本研究は JSPS 科研費 18K11456 の助成を受けたものである。

参考文献

- [1] 白畑知彦. 英語指導における効果的な誤り訂正. 大修館書店, 東京, 2015.
- [2] Takumi Gotou, Ryo Nagata, Masato Mita, and Kazuaki Hanawa. Taking the correction difficulty into account in grammatical error correction evaluation. In **Proc. of 28th International Conference on Computational Linguistics**, pp. 2085–2095, 2020.
- [3] Nitin Madnani, Martin Chodorow, Joel Tetreault, and Alla Rozovskaya. They can help: Using crowdsourcing to improve the evaluation of grammatical error detection systems. In **Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies**, pp. 508–513, 2011.
- [4] Hwee Tou Ng, Siew Mei Wu, Yuanbin Wu, Christian Hadwinoto, and Joel Tetreault. The CoNLL-2013 shared task on grammatical error correction. In **Proc. 17th Conference on Computational Natural Language Learning: Shared Task**, pp. 1–12, 2013.
- [5] Shinichiro Ishikawa. **A new horizon in learner corpus studies: The aim of the ICNALE project**, pp. 3–11. University of Strathclyde Publishing, Glasgow, 2011.
- [6] Ryo Nagata, Edward Whittaker, and Vera Sheinman. Creating a manually error-tagged and shallow-parsed learner corpus. In **Proc. of 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies**, pp. 1210–1219, 2011.

A 作例の詳細

内省に基いて作成した作例の詳細を示す。表 1 では、作例を誤りのカテゴリ（動詞、限定詞／数、前置詞）に区分けし、直感に基いて訂正の重要度順に並べている。

「動詞」では、**listened to** と **heard** とで表す意味が異なるため重要度が一番高いとした。また、二行目は動詞の意味は同じであるが、現在／過去の違いを考慮し二番目に高いとした。最後の例は、表す意味はほぼ同じと予想し、一番重要度が低いとした。

「限定詞／数」における最初の二例は、訂正前後で伝わる意味の差異は同じ程度と考えられるが、表の上では同順位が表せないため、便宜上この順で表している。一方、最後の例は a と an の使い分けに関する誤りであり、伝わる意味にほとんど差異がない。したがって、先の二例より重要度は低いとした。

「前置詞」における最初の例は、**about**→**φ**では「知っている」という基本の意味は共通するが訂正前後でニュアンスが異なるため最も重要度が高いとした。残り二例は最初の例と異なり文脈や意図とは関係なく常に誤りである。二つ目の例 **agrees φ**→**to** は、**to** 以外にも訂正候補が考えられる。一方、三つ目の例 “We discussed **about**→**φ** the issue.” については（前置詞の訂正としては）**about** を削除する方法以外は訂正候補がない。言い換えれば、前後の単語から **about** が必要がないことがほぼ確定することから、このような重要度順とした。

B ICNALE における結果の詳細

表 4 に、ICNALE を対象にして訂正の重要度を定量化した下位 10 件を示す。ただし、独自に訂正情

表 4 ICNALE を対象にした訂正重要度（下位 10 件）。

重要度	訂正文
0.01	Smoking is a common term used to involve in a habit of inhaling tobacco based products like cigarette, cigar → cigars etc..
0.01	... , surely there must be consideration before because working and learning are the side that is → are difficult to put together.
0.01	if a young man does not fall into bad habits and work hard and have → has healthy activities ...
0	... , but not limited to stroke, heart disease → diseases , brain damage, and cancer.
0	... when both college and part-time job workload takes its → their toll on the student.
0	To conclude, although there still exists a controversy, we all know that smoking is harmful for → to all people.
0	A student is the future of the nation, if we do not pay attention on → to this problem, we will lost our future.
0	Some people maybe do smoking after they finish their eat → eating , but how about other people ?
0	Weighing the two kinds of views, I think it is beneficial for us to try part-time jobs but not putting all the time and energy on → in it.
0	... because the smoke composition can help the effect of greenhouse effect that will warm the earth temperatures → temperature .