人手及び自動校閲による学習者英作文の校閲量と校閲内容 : ICNALE Edited Essaysのデータを用いて

石川, 慎一郎

(Citation)
Journal of Corpus-based Lexicology Studies,2:31-49

(Issue Date)
2020-03-15

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/81011990

(URL)
https://hdl.handle.net/20.500.14094/81011990

英語コーパス学会 語彙研究会 (Japan Association for English Corpus Studies, Lexicology SIG)
Journal of Corpus-based Lexicology Studies, Vol. 2 (March, 2020)
ISSN: 2434-169X

人手及び自動校閲による学習者英作文の校閲量と校閲内容

―ICNALE Edited Essays のデータを用いて―

石川　慎一郎（神戸大学）
iskwshin@gamail.com

Human and Automated Editing of L2 English Learner Essays
―Based on the Data from the ICNALE Edited Essays―

ISHIKAWA Shin'ichiro (Kobe University)

概要
ユーザーの英文中の誤りを検出し，修正候補を提案する英文自動校閲システムが広く使用されるようになっている。こうしたシステムは学習者にとっては利便性が高いが，校閲の量や内容がどの程度妥当なものであるのか，人手による校閲結果とどの程度一致しているのか，といった点については必ずしもはっきりしていない。本研究では，アジア圏国際英語学習者コーパス ICNALE に含まれる学習者オリジナル作文と人手校閲作文に加え，代表的な自動校閲システムで生成した自動校閲作文を比較することで，人手校閲と自動校閲の特性を検証した。

キーワード
英文自動校閲，学習者コーパス，校閲量，校閲内容

1. はじめに

近年，オンライン上で，または，ワードプロセッサソフトなどに組み込む形で，手軽に使用できる英文自動校閲システムが多数開発され，幅広いユーザーに使用されている。これらは，英作文を苦手とする学習者はもちろんのこと，学習者作文の添削や指導に多くの時間を費やしている英語教師にとっても潜在的に有益なツールとなりうる。しかし，こうしたシステムによって，作文中の誤用がどの程度検出され，それらに対してどの程度適切な修正指示がなされているかは必ずしもはっきりしない。

そこで，簡易な実験として，誤用文（*I think I and you am a frend.）を作成し，5 種の自動校閲システムで処理を行ったところ，以下の結果を得た（調査は 2020 年 1 月に実施）。

表 1 5 種の校閲システムによる自動校閲結果の比較

校閲システム名	自動校閲後の英文（（筆頭）修正候補をそのまま採用）
(A) Grammarly	I think you and I am a friend.

ISSN: 2434-169X

(B) Pro Writing Aid	I think I and you am a friend.
(C)After the Deadline	I think I and you am a friend.
(D) 1 checker	I think I and you as a friend.
(E) Ginger	I think I and you am a friend.

当該英文には 5 つの誤用が含まれていたが，(1)語順の誤り(*I and you → you and I)は，(A)を除く 4 種で検出されなかった。(2)主語と動詞の一致の誤り(*am→are)は，(D)を除く 4 種で検出されなかった。また，(D)も am を are ではなく as に置き換える不適切な修正を行っていた。(3)相互複数の誤り(*friend→friends)は，5 種すべてで検出されなかった。(4)単語のスペルの誤り(*frend→friend)のみ，5 種すべてで検出され，適切に修正されていた。この小規模な実験結果に即して言うと，自動校閲システムによる誤りの検出・修正には問題が多いように思える。

もっとも，上記は人為的に作成した短い一文の分析結果にすぎず，一般の英語学習者が自然に産出した一定量の長さのテキストについて自動校閲システムがどのような校閲結果を返すのか，それが人手による校閲結果とどの程度一致しているのかは明らかでない。そこで本研究では，学習者オリジナル作文，人手校閲作文，自動校閲作文の 3 種を用意し，それらを相互に比較することで，人手校閲と自動校閲による校閲の量と内容を比較することとしたい。なお，自動校閲システムについては，類似のシステムの中でとくに完成度が高く，広く使用されている Grammarly を検証対象とする。なお，Grammarly を含む各種の自動校閲システムは頻繁にアルゴリズムの改良を行っており，本稿で示す結果は当該システムの恒常的・絶対的な性能を示すものではない。

2. 先行研究

Grammarly を含め，多くの自動校閲システムは企業内の研究成果をふまえて開発・更新されており，誤り検出や修正候補の選択アルゴリズムが一般に公開されることはほとんどない。しかし，こうしたシステムの開発には，工学(自然言語処理)分野における「文法誤り訂正(Grammar Error Correction：GEC)」研究の成果が活かされていると考えられる。

浅野・水本・乾(2017)によれば，自然言語処理の分野では，いわゆる"shared task"として，GEC の精度を競うコンペティションが 2011 年から 4 年連続で実施され，これにより，GEC 研究のすそ野が世界に広がったとされる。こうした動きは現在も継続しており，たとえば，2019 年には，ケンブリッジ大学によって，Building Educational Applications 2019 Shared Task: Grammatical Error Correction と名付けられたコンペティションも実施されている(https://www.cl.cam.ac.uk/research/nl/bea2019st/)。

こうしたコンペティションでは，共通の問題と正解(レファレンス)が用意され，研究者は独自に開発した誤り検出手法の性能を競い合う。手法の評価では，課題英文に存在する誤りのうちいくつを見つけたかという検出の網羅性(再現率：recall)と，指摘した誤りのうちいくつが実際に誤りであったかという検出の精度(適合率：precision)が重要であり，通例，手法の性能は F 値(適合率と再現率の積を 2 倍して，適合率と再現率の和で割る)で示される。

一般に，誤り検出を自動化しようとする場合，英作文における誤りの分布実態を把握する必要がある。甲南大学と教育測定研究所（JIEM）が開発した Konan-JIEM Learner Corpus（3 版）に付されたエラータグによると，品詞別に見た誤りは，動詞（24%）＞名詞（22%）＞冠詞（19%）＞前置詞（13%）とされる（牧野・新妻・太田，2016）。また，Wordvice（2016）は，同社が受注した数百万語の英文校正データを解析し，タイプ別に見た誤りは，スタイル（32%：受動態使用・冗長表現使用など）＞語選択（22%）＞文法（21%：冠詞・前置詞・主語の数と動詞形の呼応の誤りなど）＞スペル（14%）＞句読法（9%：不適切なコンマによる文要素の分離など）＞文構造（2%：形容詞の位置誤り・断片文など）で，文法誤りの 62%が冠詞の誤りであるとしている。

これらをふまえ，工学系の研究者たちは，特定のタイプの誤りを取り上げ，それを最も効率的に検出する手法の開発を行ってきた。とくに，明確な用法ルールの存在しない冠詞と，意味が多様で誤りの検出が難しい前置詞をターゲットとした研究が多い。以下，最近の国内の研究の一端を概観する。

まず，冠詞の誤り検出に関するものとして，若菜・永田・桝井・河合（2005）は，名詞の周辺単語情報に基づいて，名詞の可算・不可算判定を行う手法を開発した。検証の結果，既存システムの再現率が 0.13 であるのに対し，提案手法では 0.59 を達成したとされる。また，乙武・荒木（2010）は，従来型の最大エントロピー分類法と，提案手法である意味カテゴリ情報を用いた帰納学習法の性能を比較した結果，前者で無冠詞かどうかの判定を行い，後者で冠詞の選定を行う組み合わせ処理が望ましいと述べている。

次に，前置詞の誤り検出に関して，久保田・太田（2012）は，調べたい前置詞を含む英文を入力すると，そこから検索フレーズを自動生成してウェブ検索を実行し，その中の前置詞頻度に基づき修正候補を抽出するシステムを開発した。評価実験の結果，置換誤り検出の F 値は 0.896，挿入誤り検出の F 値は 0.80 であったとしている。また，永田・水本・Whittaker（2013）は，前置詞の誤り検出を，特定の前置詞の正誤を判定する 2 値分類，あるいは，正しい前置詞を選択する n 値分類としてとらえる従来型のアプローチでは，誤りの検出・訂正ができたとしても，なぜその前置詞が間違っているかの説明はできず，教育的汎用性に欠けると指摘した。その上で，解釈が容易で，言語知識を反映でき，学習者へのフィードバックを付与できる「誤り格フレーム」という情報記述の枠組みを提案し，母語話者コーパスと学習者コーパスからこうしたフレームを自動生成する手法の開発を行った。自動生成されたフレームの妥当性を人手で検証した結果，日本人学習者の作文データの場合は，0.707 の精度が得られたと言う。また，前置詞誤りのうち，脱落に関するものが 4 割を占めることも示された。

このほか，牧野・新妻・太田（2016）は，名詞の選択誤り（sunlight とすべきところを sun とするなど）を近傍形容詞と当該名詞の共起強度から自動検出する手法を開発し，性能検証の結果，F 値は 0.53 であったという。松井・新妻・太田（2016）は，動詞の時制誤りに関して，これを，動詞に正しい時制ラベルを割り当てる「系列ラベリング問題」とみなした上で，文脈情報を組み込んだ検出手法の開発を行い，検証の結果，$F_{0.5}$ 値が 0.329～0.360 であったとしている。また，判定に有効な素性が動詞見出し語・副詞・等位接続詞であったことも明らかになった。谷本・太田（2011）は，コロ

ケーション誤り（have a dream を see a dream とするなど）に関して，検討したい英文を入力すると，検索クエリが自動で生成され，ウェブ検索数により，コロケーションの適否を判定し，あわせて修正候補を提示する仕組みを開発し，既存手法を上回る性能を達成したとしている。

このように，「文法誤り訂正」に関連する工学系の研究は多岐にわたっている。内外での同様の研究で公開された各種の知見が，企業の自動校閲システムにも反映されているものと推定される。

3. リサーチデザイン

3.1 研究目的

本研究の目的は，人手校閲と自動校閲による校閲の量と内容を計量的に比較し，英文自動校閲システムの信頼性を検証するとともに，教育現場における当該システムの望ましい使用方法に関して一定の指針を得ることである。この目的に沿い，以下の 4 つのリサーチクエスチョンを設定した。

RQ1：人手校閲と自動校閲の全体校閲量はどのような関係にあるか？
RQ2：人手校閲と自動校閲の観点別校閲量はどのような関係にあるか？
RQ3：書き手の習熟度と人手校閲の校閲量及び自動校閲の観点別校閲量はどのような関係にあるか？
RQ4：人手校閲と自動校閲の内容にはどのような特徴があるか？

3.2 データ

本研究では，オリジナル作文，人手校閲作文，自動校閲作文を比較する。以下，本研究で使用したデータの概要について述べる。

3.2.1 オリジナル作文・人手校閲作文

本研究では，世界最大級のアジア圏英語学習者コーパスである International Corpus Network of Asian Learners of English（ICNALE）のデータを分析サンプルとして利用する。ICNALE には，アジアの EFL 圏 6 地域（CHN：中国，IDN：インドネシア，JPN：日本，KOR：韓国，THA：タイ，TWN：台湾）と ESL 圏 4 地域（HKG：香港，PAK：パキスタン，PHL：フィリピン，SIN/MYS：シンガポール及びマレーシア）の大学生（大学院生含む）と，英語母語話者による英語の産出データが収録されている。

ICNALE は，作文を集めた Written Essays（2,800 人／130 万語）（Ishikawa, 2013），作文評価データと校閲済作文を集めた Edited Essays（320 人／15 万語）（Ishikawa, 2018），モノローグ発話を集めた Spoken Monologue（1,100 人／50 万語）（Ishikawa, 2014），インタビューでのダイアローグ発話を集めた Spoken Dialogue（425 人／160 万語）（Ishikawa, 2019）の 4 つのモジュールで構成されているが，今回使用したのは，Edited Essays のデータである。

ICNALE Edited Essays には，Written Essays からサンプリングされた合計 640 本のオリジナル作文と，専門の校閲者による観点別作文評価データ，また，校閲済みの作文が収録されてい

る。オリジナル作文のテーマは,「大学生のアルバイトの是非」(It is important for college students to have a part-time job)と「レストランでの禁煙の是非」(Smoking should be completely banned at all the restaurants in the country)の 2 種で, それぞれ 20～40 分間で 200～300 語の作文を行った。辞書の使用は禁じたが, ワードプロセッサに付属するスペルチェッカの使用は義務付けた。

校閲者は 5 名で, いずれも, 学術論文の校閲を, 2 年以上, 職業として行っている母語話者である。校閲に先立ち, 研修を実施し, 同じ水準での校閲が提供されるよう配慮した。校閲者には, オリジナル作文の内容を変更しないことを義務付けた。一方, 英語については, 当該作文を読むだけで意味が読み取れる("fully intelligible")レベルに校閲するよう求めた。作業は, MS Word の校閲機能を用いて実施させたため, 遡って, 加除の数や内容が特定できるようになっている。

今回は, 日本人大学生 40 名が「レストランでの禁煙の是非」について書いた 40 本のオリジナル作文と, 同数の人手校閲済み作文を分析対象とする。なお, これらはすべて同一の校閲者によって作業されたもので, 校閲者の違いによる影響はない。また, 学習者は, TOEIC, TOEFL などの英語習熟度テストのスコア, または語彙サイズテストのスコアによって, ヨーロッパ言語共通参照枠(Common European Framework of Reference for Languages:CEFR)に基づく A2, B1_1, B1_2, B2+の 4 レベルに分類されている。TOEIC L/R テストのスコアで言えば, A2 は 545 点未満, B1_1 は 550 点以上, B1_2 は 670 点以上, B2+は 785 点以上に相当する。

3.2.2 自動校閲作文

今回は, 各種の自動校閲システムの中でも使用者が多い Grammarly を選び, 40 本のオリジナル作文について自動校閲版を作成した。Grammarly には様々なバージョンがあるが, 今回使用したのは, MS Word に組み込んで使用する Grammarly for Microsoft® Office である。アカウントは個人向け有償版(Premium)で, 誤りの検出対象や修正指示が無償版より拡充されている。

Grammarly では, 文章のタイプや校閲の対象を指定することができる。前者については, 一般・学術・ビジネス・技術・医療・創作・カジュアルの 7 種から選べるが, ここでは「一般」を指定した。また, 後者についても, スペル(contextual spelling), 文法(grammar), 句読法(punctuation), 文構造(sentence structure), 文体(style), 語彙の言い換え(vocabulary enhancement), 盗用(plagiarism)の 7 種が用意されているが, ここでは, 人手校閲の条件とそろえるため, スペル・文法・句読法・文構造の 4 種のみを指定した。

誤りを検出した場合, Grammarly では修正候補が別画面に表示され, ユーザーの選択によって修正が実行される。 本来はユーザーが指摘の内容を吟味して反映・非反映を決定する仕組みであるが, 今回は, 初級学習者による自動校閲システムの使用実態をふまえ, 内容の吟味を行わず, 機械的に(筆頭)修正候補を反映することとした。なお, センテンスが不完全である場合など, 問題がありうることだけが指摘され, 修正案が示されないことがあるが, この場合は修正を行っていない。

3.3 手法

RQ1 に関して，人手校閲は，MS Word の「変更履歴」を確認し，挿入箇所と削除箇所の合計を全体校閲量とみなした。図 1 の例では 17＋9=26 となる。また，自動校閲は，4 観点で検出された問題箇所の合計値を 2 倍したものを全体校閲量とみなした。図 2 の例では，(2+15+3+1)*2=44 となる。問題箇所の処理には，削除のみ，挿入のみ，削除＋挿入の 3 つのパタンがありうるが，ここでは，最大推定として，問題箇所のすべてで削除＋挿入の 2 ステップが必要であったと仮定している。これにより，挿入と削除の数を積み上げて校閲量とみなした人手校閲の場合とほぼ等しい条件で比較ができる。

図 1 人手校閲の校閲量特定

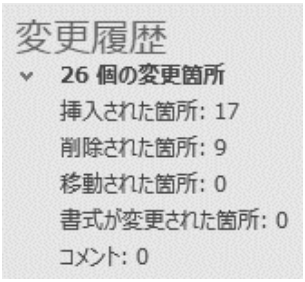
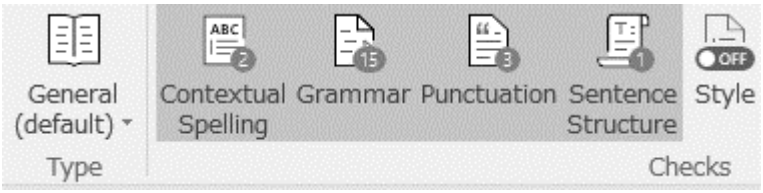


図 2 Grammarly の校閲量特定



以上の前提のもと，40 本の作文の各々について人手校閲と自動校閲の校閲量を求め，その平均値を比較する。さらに，ピアソンの積率相関係数により，両者の関係性を検証する。

次に，RQ2 では，Grammarly のタイプ別校閲量のデータを用い，RQ1 と同様の調査を行う。上記の例で言えば，スペル・文法・句読法・文構造の校閲量は，それぞれ，2*2=4，15*2=30，3*2=6，1*2=2 となる。

RQ3 では，各種の校閲量を第 1 アイテム(人手:挿入，人手:削除，人手:全体，自動:スペル，自動:文法，自動:句読法，自動:文構造，自動:全体の 8 カテゴリ)，40 本の作文を第 2 アイテム(40 カテゴリ)とする頻度表を作成し，対応分析(correspondence analysis)を実施する。なお，第 2 アイテムの検体名は学習者の習熟度レベル(A2, B1_1, B1_2, B2)とする。対応分析では，分割表の行列間の相関を最大化するよう複数の次元が取り出されるが，一般的には，寄与率の最も高い 2 つの次元を横軸・縦軸として散布図を作成し，第 1 アイテム・第 2 アイテムを散布図上に同時布置する。これを質的に解釈することにより，データの分布特性を調べることができる。

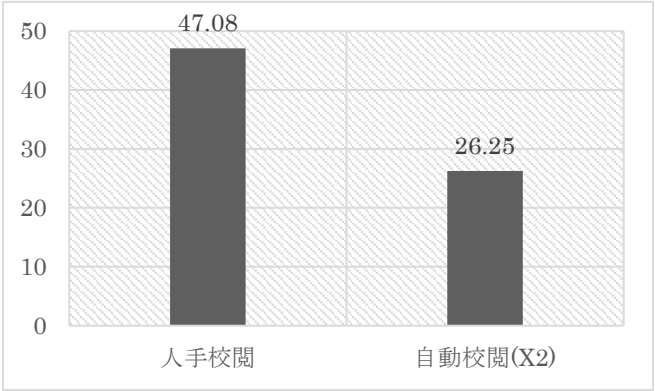
最後に，RQ4 では，オリジナル作文と 2 種の校閲作文に出現する全単語の頻度を対数尤度比統計量(log-likelihood ratio:G^2)で比較し，統計量の上位 20 語(同一統計量に複数語が並ぶ場合はランダムに選定)を抽出する。オリジナル作文側で顕著に頻出している語は，校閲によって削除された語であり，校閲作文側で顕著に頻出している語は校閲によって新規に挿入された語と言える。なお，統計量による特徴語抽出に際しては統計量や効果量の閾値を設定することもあるが，今回は，幅広い語を観察できるよう，閾値は設定しない。

4. 結果

4.1 RQ1　全体校閲量

まず，人手校閲による校閲量と自動校閲による全体校閲量を比較したところ，以下の結果を得た。

図 3　人手校閲と自動校閲の校閲量



t 検定（Welch）の結果，両者の平均値の差は有意であり（t=6.01: df=71.59; p< .001），人手校閲に比べると，自動校閲で検出できる誤りの箇所は少ないことが確認された。今回，自動校閲については，問題点が指摘された箇所のすべてで削除・挿入が発生しているとみなす最大推定を行ったわけだが，それでも校閲量は人手校閲より 4 割ほど少ないという結果となった。

では，2 種類の校閲量はどの程度相関しているのであろうか。人手校閲と自動校閲がほぼ同様のロジックで誤り検出を行っているのであれば，絶対数に差があっても，個々の作文に対する校閲量の多寡はそろうはずで，つまり，相関係数は限りなく+1 に近くづくはずである。この点をふまえ，40 本の作文に対する相関を調べたところ，相関係数は r=.687 であった。説明力は約 48%にとどまり，2 種の校閲による校閲量の相関は予想よりもかなり限定的であることがわかった。このことは，自動校閲が人手校閲とはかなり異なるロジックによって誤りの検出を行っている可能性を示す。

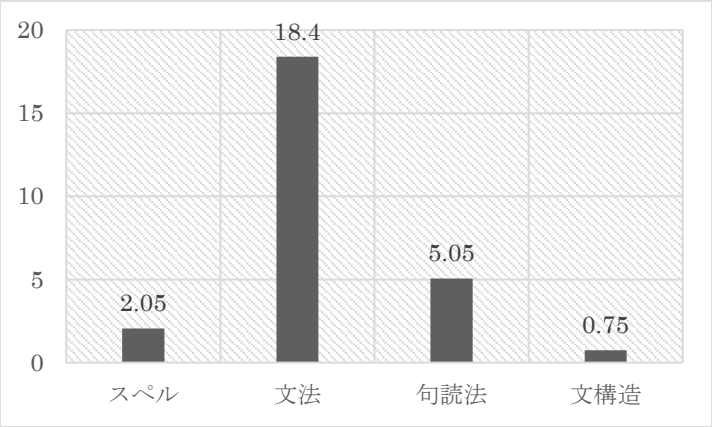
4.2 RQ2　自動校閲の観点別校閲量

自動校閲の校閲量を 4 観点にわけて調査したところ，図 4 の結果を得た。図からわかるように，文法の校閲量が他を圧倒している。

分散分析の結果，校閲観点の違いが校閲量に及ぼす効果は有意で（*Mauchly's W*=0.18; *Chi^2*=65.62; df=5; p< .001），Bonferroni 法による多重比較の結果，スペルと文構造間を除き（p=1.00），観点間の差も有意であることが確認された（スペルと句読法間は p<.05，他はすべて p<.001）。

つまり，校閲量は，文法が圧倒的に多く，以下，句読法＞スペル≒文構造の順となっていることになる。これは，スタイルと語選択を除くと，英文中での校閲箇所が，文法＞スペル＞句読法＞文構造であったとする Wordvice（2016）の報告内容とほぼ一致している。

図 4 自動校閲の観点別校閲量



実際の校閲内容を概観したころ，文法については，動詞の活用形ミス，冠詞の漏れ（または間違い），名詞の単複の不一致などへの指摘が多かった。次に，句読法については，センテンス内の要素の切れ目にコンマを挿入するよう指示する例が多かった（文 1 and 文 2→文 1, and 文 2, など）。スペルについては，オリジナル作文の提出時にスペルチェックを行わせているわけだが，チェック漏れによる綴り誤りや，語の派生形の誤り（Japan restaurants→Japanese restaurants），また，複合語の表記誤り（non smoker → nonsmoker/ non-smoker）や縮約形の表記誤り（don‘t → don’t）の修正が見られた。文構造に関しては，否定辞の文内位置の誤り（will be not able to → will not be able to）などについての指摘が認められた。

続いて，自動校閲による 4 観点別及び全体校閲量，また，人手校閲の校閲量間の相関を調べたところ，表 2 の結果を得た。

表 2 各校閲量間の相関

	自動校閲（X2）					人手校閲
	スペル	文法	句読法	文構造	全体	
スペル	1.000	***	*n.s.*	*n.s.*	***	**
文法	0.536	1.000	*n.s.*	*n.s.*	***	***
句読法	0.154	0.210	1.000	*n.s.*	**	*
文構造	0.013	0.062	0.192	1.000	*n.s.*	*n.s.*
全体	0.716	0.929	0.449	0.190	1.000	***
人手校閲	0.472	0.612	0.373	0.231	0.687	1.000

注：上三角は母相関係数の無相関検定の結果を，下三角は相関係数を示す。

まず，自動校閲の観点別校閲量と全体校閲量に注目すると，文法（r=.929）とスペル（r=.716）は全体相関量に強く相関していたが，句読法の相関は中程度（r=.449），文構造の相関は弱い（r=.190）ことがわかった。

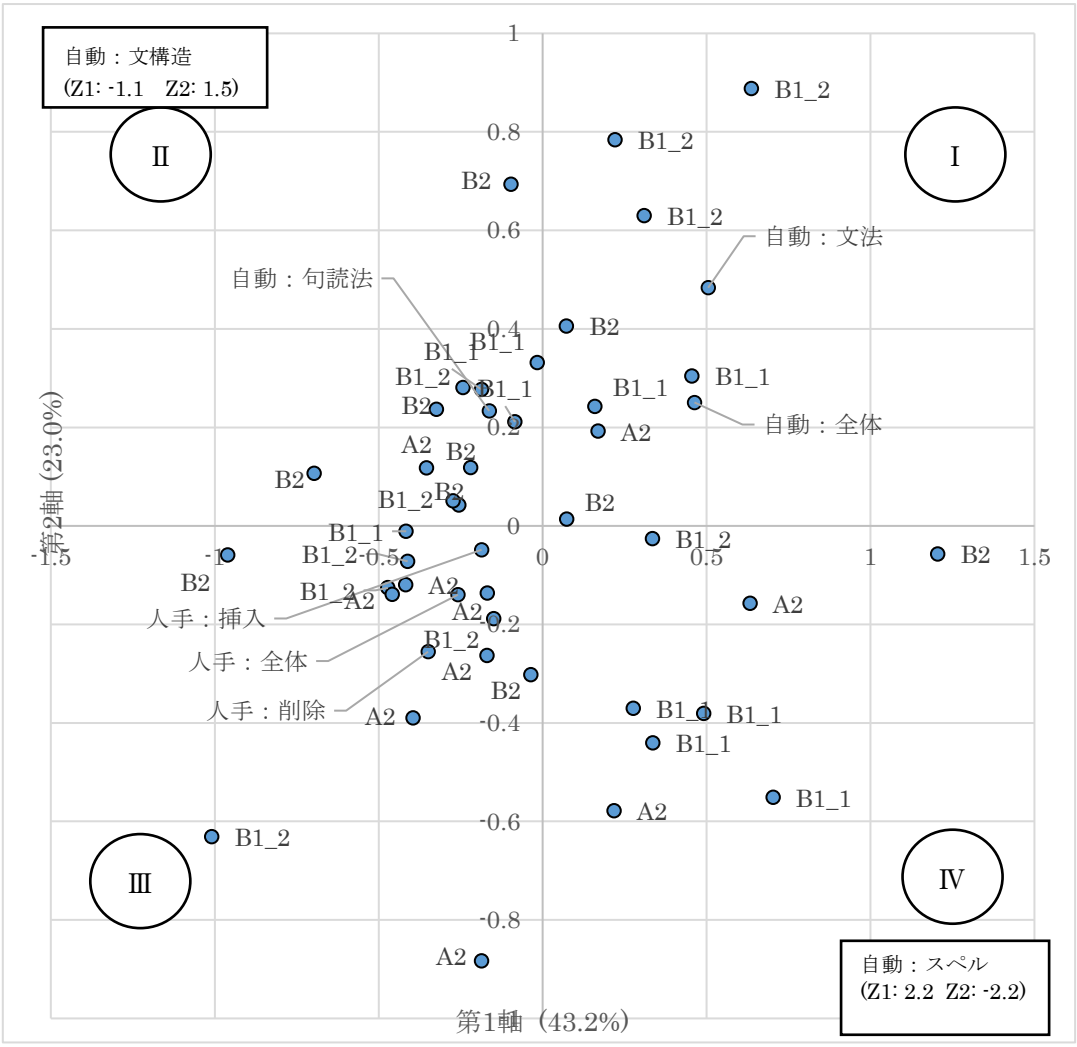
次に，観点別の相関に注目すると，スペルと文法間には中程度の相関が認められたが（r=.536），他の組み合わせでは相関は弱く，観点間のオーバーラップは限定的であることが確認された。

最後に，人手校閲の校閲量との相関は，全体（r=.687）＞文法（r=.612）＞スペル（r=.472）＞句読法（r=.373）＞文構造（r=.231）となった。自動校閲の全体校閲量及び文法校閲量は人手校閲量とある程度関係しているが，その他の観点については自動校閲側の独自の判断によってなされている可能性が示された。

4.3 RQ3　書き手の習熟度との関係

続いて，対応分析を行ったところ，以下の散布図を得た（図 5）。 なお，自動校閲による文構造の校閲量と，同じく自動校閲によるスペルの校閲量は，散布図の中で，他と大きく離れた位置に布置されていたため，下図ではそれらについては別途書き加えている。

図 5 対応分析に基づく散布図



まず，全体は第1軸（分散の 43.2%を説明）によって左右に分割される。右側には，A2 や B2 の混入がいくつかあるものの，全体として見れば，B1（B1_1 と B1_2）が多く布置されている（右側に布置された 16 の学習者作文のうち，B1 は 10 種（63%）を占める）。一方，左側には B2 と A2 が多く布置されている（左側に布置された 24 の学習者作文のうち，B2 と A2 は 14 種（58%）を占める）。これより，横軸は B1 とそれ以外を区分する軸になっていると考えられる。B1 レベルの学習者の作文は他のレベルの学習者の作文に比して，誤用という点で，一定のまとまりと独自性を持っているようである。

一方，縦軸に注目すると，上方には B2 や B1_2 など高習熟度の学習者が，下方には A2 や B1_1 など低習熟度の学習者が布置されており，縦軸は，下方から上方にかけて，全般的な習熟度の上昇を表す軸になっていると思われる。

以上の 2 つの軸を組み合わせると，全体は 4 つの象限に区分できる。各象限の特徴は表 3 のようにまとめられる。

表 3 対応分析に基づくグルーピング

象限	学習者習熟度	対応する校閲タイプ
Ⅰ	B1_2 など	自動校閲（文法），自動校閲（全体）
Ⅱ	B2 など	自動校閲（文構造），自動校閲（句読法）
Ⅲ	A2 など	人手校閲（挿入・削除・全体）
Ⅳ	B1_1 など	自動校閲（スペル）

上記に基づくと，B1 レベル以上の学習者の作文は自動校閲でも一定の対応が可能であるが，A2 レベルの初級者の作文については自動校閲ではうまく対応できず，人手による校閲が必要であると言えそうである。

4.4 RQ4 校閲内容

最後に，校閲前後の語の頻度を比較することで，人手校閲と自動校閲における特徴的な削除・挿入語を調査したところ，表 4 の結果を得た。それぞれ，上位 20 語を選んだ場合の対数尤度比統計量（G^2）の下限を示し，あわせて，5%有意水準に相当する G^2=3.84 を満たしている語に下線を付している。

表 4 人手校閲と自動校閲における特徴的校閲内容（削除語・挿入語）

人手校閲	
削除語（校閲前＞校閲後）G^2 >=1.3	挿入語（校閲前＜校閲後） G^2 >=4.0
smoker, person, restaurant, area, things, people, so, section, man, something, part, called, enter, its, less, place, from, room,	smokers, secondhand, areas, manners, enjoyment, firsthand, develop, establishing, fewer, fully, impacted,

seat, every	their, lung, seats, would, smells, bodies, cafés, elderly, evidence
自動校閲	
削除語(校閲前＞校閲後)G^2>=0.2	挿入語(校閲前＜校閲後)G^2>=1.4
smoker, chose, cigarette, less, non, ask, break, comparing, include, let, look, tastes, theirs, wants, year, which, area, place, body, consider	a, nonsmoker, the, against, dividing, evidence, fewer, staff, another, asks, bodies, breakfast, compared, curtains, feelings, healthily, heart, includes, letting, lives

はじめに注目すべきは，対数尤度比の下限値で，削除の場合も，挿入の場合も，統計値は自動校閲のほうが小さくなっている。このことは，人手校閲が 40 本の作文を通して同じ語を何度も繰り返して削除したり挿入したりしているのに対し，自動校閲では校閲の全体量が限られていたこともあり，同一語の加除の回数が相対的に少なかったことを示す。また，それぞれの校閲において，典型的な修正のパタンがあることもわかった。

まず，人手校閲では，以下のような修正の傾向性が認められた。

(a) 単数名詞の複数化(smoker→smokers, area→areas, seat→seats)
(b) 総称名詞の削除(person, man, people, things, something→ φ)
(c) 接続詞の so(だから)の削除(so→ φ)
(d) 動詞の多様化(φ →develop, impacted, enjoyment, establishing など)
(e) ヘッジ表現の追加(φ →would)

これらは，先行研究で明らかにされてきた日本人学習者の過剰使用語(石川，2012 他)ともかなり重なっている。先行研究では，日本人学習者は，名詞の数に対する配慮が希薄なために無冠詞単数を多用し，物主構文を避けるために総称的人称代名詞で文を始め，おそらくは L1 影響もあって因果関係を示す接続詞を好み，限定的な語彙を用い，ヘッジを置かずにしばしば極端な断言を行うことが指摘されてきたわけだが，これらの多くが，単なる過剰使用にとどまらず，文法的誤用として修正されていることは興味深い事実と言える。

次に，自動校閲では，以下のような修正の傾向性が見られた。

(a') 単数名詞の複数化(body→bodies)
(b') 単数名詞への冠詞付与(φ →a/ the)
(c') 名詞の可算・不可算の誤り修正(less→fewer)
(d') 動詞活用形の修正(ask→asks，comparing→compared，let→letting，include→includes)

(e') 複合語表記の変更(non smoker→nonsmoker)

このうち, (a')～(c')は, 日本人学習者の名詞の数に対する配慮の希薄さに起因するもので, 前出の(a)と対応している。また, (d')は前出の(d)とある程度関連していると言える。これに対し, (e')は自動校閲の側でのみ目だって多くなされた修正である。一方, 前出の(b), (c), (e)に関係する修正は自動校閲では多くは見られなかった。人手校閲が, ある程度, 文の内容やスタイルに踏み込んで修正を行っているのに対し, 自動校閲は(あらかじめ校閲対象から「スタイル」を除去していたこともあるが)表層の文法や表記の誤りに特化して修正を行っていることがうかがえる。

以上より, 人手校閲と自動校閲によって加除された内容には, 名詞の数に関する誤りの修正や, 動詞の選択や活用形に関する誤りの修正など, 共通する部分もあるが, 自動校閲の内容は, 冠詞の付加や語形(活用形)の変更が大部分を占め, 語全体の削除や別語への置き換えは少ないことが示された。

5. 質的考察

以上で, 4つのリサーチクエスチョンに即して, 2 種の校閲の差異を計量的観点から概観してきた。ここでは, これまでの調査で明らかになった点をふまえつつ, 2種の校閲を経て, 学習者のオリジナル作文がどのように修正されたかを質的に検証し, 各々の校閲内容について議論を深めたい。

サンプルとするのは, JPN_024 の作文である。当該学習者の英語力は初級(A2 レベル)で, 今回, 分析した 40 本の中で最も多くの校閲が加えられていた。ここでは, オリジナル作文の冒頭 7 センテンス分を考察の対象とし, 便宜上, 第 1~2 文, 第 3~4 文, 第 5~7 文の 3 つにわけて概観していく。

はじめに, 第 1~2 文を比較する。

表 5 オリジナル作文と校閲作文(第 1~2 文)

文	オリジナル	人手校閲	自動校閲
1	Yes, I agree with the statement.	I agree with the statement because I hate smoking while I am eating a meal.	Yes, I agree with the statement.
2	Because I hate smoking with eat meal.		Because I hate smoking to eat meals.

オリジナル作文には, 少なくとも, (1)冒頭の会話的な Yes, (2)because の独立節での使用, (3)meal の無冠詞単数形使用, (4)with eat meal という非文法的表現, という 4 つの問題がある。

(1)については, 人手校閲では正しく削除されているが, 自動校閲では誤りとされなかった。事前に文体を「一般」に指定していたわけだが, こうした口語的な挿入語は誤り検出の対象となっていないようである。

(2)は学習者のおかしやすい誤りで,『ジーニアス英和辞典』にも,「よく日本人は ×I'm sorry I

couldn't be there. Because I was busy. のように主節と because 節を別々に書くが，これは誤り」であるという明快な注記がある。人手校閲は because を第 1 文に接合して問題を解消しているが，自動校閲ではこの点も誤りとされなかった。

(3)も学習者のおかしやすい誤りであるが，人手校閲は不定冠詞を付すことで，自動校閲は複数化することで，どちらも文法的な問題は解消されている。ただし，今回の例では，個人の 1 回の食事時間中の喫煙を問題にしているわけなので，meals とした自動翻訳よりも，a meal とした人手翻訳のほうがより妥当であろう。

(4)は，前置詞 with と動詞 eat が結合するという文法的な誤りを含むだけでなく，そもそも，eat と前節する smoking の動作主がはっきりしないという意味論的な問題も抱えている。人手修正は while I am eating a meal とすることで，文法上の問題は解消しているが，smoking の動作主ははっきりしない。明示されていない動詞の動作主は主節の動作主に従うとすれば，この箇所の意味は「自分は食事中にタバコを吸うのは嫌なのでこの意見（レストランでは禁煙すべき）に同意する」となるが，作文の後半で書き手は非喫煙者であることが述べられているので，これでは意味が通らない。一方，自動校閲は， with eat meal を to eat meals に置き換えることで，これも文法上の問題は解消しているが，「食事をするために喫煙を嫌う」という趣旨になってやはり意味は通らない。書き手の意図を汲み取って解釈すれば，この箇所は「自分が食事をしているときに《他の人》がタバコを吸っているのは嫌だ」といった内容になるのが自然で，英語としては，I do not want others to smoke while having a meal.のように修正すべきであったと考えられる。

次に，第 3~4 文を見てみよう。

表 6 オリジナル作文と校閲作文（第 3~4 文）

文	オリジナル	人手校閲	自動校閲
3	Smoker has right smoking but non smoker also has right not smoking at the same time.	Smokers have the right to smoke, but non-smokers also have the right not to smoke at the same time.	The smoker has the right smoking, but nonsmoker also has right not smoked at the same time.
4	But non smoker's right is lighter than smoker's one.	But the non-smoker's right is lighter than smoker's.	But nonsmoker's right is lighter than the smoker's one.

オリジナル作文には，少なくとも，(1')冒頭の大文字 But，(2')単数名詞の無冠詞使用（(non) smoker, right），(3')複合形の表記誤り（non smoker），(4')非文法的な right smoking 及び right not smoking，(5')所有格の後の代名詞 one，(6')不自然な right is light というコロケーション，の 6 つの問題点が認められる。

(1')については，人手校閲・自動校閲とも，誤りとみなしていない。伝統的な英作文指導では，文

頭の大文字の And や But は避けるよう教えてきたため，これは意外な結果に見えるが，最近では，大文字の And や But を問題視することは減っているようである。『ジーニアス英和辞典』も，「従来は，書き言葉で but を文頭に用いるのは避けた方がよいとされてきたが，新聞・雑誌・小説などの分野では確立した語法になってきている」という注記を示している。この点をふまえれば，But を誤りとしていない人手校閲・自動校閲の判断は一応理解できる。

(2')について，人手校閲は第 3 文の smoker→smokers, right→the right, non smoker→non-smokers や，第 4 文の non smoker's right→the non-smoker's right に見られるように，無冠詞単数が適切に複数化ないし定冠詞化されている。ただし，第 4 文の smoker's one は無冠詞のままで誤りとして検出されていない。一方，自動校閲の場合，3 文冒頭の right→the right のように正しく修正されている場合もあるが，初出の冒頭の smoker を the smoker に変えるといった不自然な修正があるほか，3 文後半の nonsmoker や right，4 文の nonsmoker's right などは誤り検出から漏れている。人手校閲も完全ではないが，自動校閲による誤り修正はより限定的であるように思われる。

(3')については，オリジナル作文で開放複合形（open compound）のように表記されていた non smoker(s)が，人手校閲ではハイフンつき複合形（hyphenated compound）に，自動校閲では連結複合形（solid compound）に修正されている。これらはいずれも妥当な修正である。

(4')について，人手校閲は right smoking→the right to smoke, right not smoking→the right not to smoke と修正しており，文法的な問題点は解消されている。一方，自動校閲は right smoking は（the を付けたほかは）そのままにしている。また，right not smoking はなぜか right not smoked に修正され，意味が通らなくなっている。なお，この箇所には，気付かれにくいが，もう 1 つの意味論的な問題が隠れている。ここで，書き手は「喫煙者は吸う権利を，非喫煙者は吸わない権利を持つ」と書いているが，続く文で，「非喫煙者の権利が軽くあしらわれている」と述べていることをふまえると，非喫煙者が持つのは「吸わない権利」ではなく，「吸わせない権利」，言い換えれば，「タバコの煙に邪魔されない権利」であることがわかる（喫煙者が吸おうが吸うまいが，非喫煙者が自分で吸わない限り，非喫煙者の吸わない権利は侵害されず保たれているため）。書き手の意図をここまで汲み取ると，この箇所の修正は，本来，non-smokers have the right to be free from the smell of smoke などとなるべきであったと思われる。

(5')については，所有格＋one は非文法的である。Mike's one や his one が言えないとの同じく，smoker's one も言えない。人手校閲では適切な修正されているが，自動校閲ではこの誤用も見逃されている。

最後に，(6')について，書き手は，「権利が軽い」という日本語からの連想で形容詞 light(er)を使ったものと思われるが，right と light のコロケーションは英語として一般的ではない。超大型の英語コーパスである enTenTen（English Web2015）で right を修飾する主な形容詞を確認したが，human, civil, fundamental, constitutional, equal, basic, intellectual, legal などはあるが，上位共起形容詞トップ 100 語中に light は含まれていなかった。英語としての自然さを重視すれば，この箇所は，the non-smoker's right is more limited などと修正されるべきであると思われ

るが，人手校閲・自動校閲とも，この点については誤りとしてしない。

最後に，第 5~7 文を概観しよう。

表 7 オリジナル作文と校閲作文（第 5~7 文）

文	オリジナル	人手校閲	自動校閲
5	Smoker's can smoke almost every where every time.	Smokers can smoke almost everywhere at any time.	Smokers can smoke almost everywhere every time, although smoker can control their desire for a short time like class.
6	Despite smoker can control their desire short time like class.	Despite the fact that smokers can control their desire for a short time like during class, if they want to smoke, they can go to a smoking place soon after eating a meal.	
7	If they wanted to smoke, they can go to smoking place after eating meal soon.		If they wanted to smoke, they could go to a smoking place after eating a meal soon.

ここにも問題点は多いが，名詞の無冠詞単数使用（smoker, meal）や複合語の綴り（every where）についてはすでに言及したので，以下では，(1”)助動詞の不適切使用（can），(2”) 不正確な慣用表現（every where every time），(3”)時制エラー（wanted），(4”)despite…の独立節の問題，の 4 点にしぼって議論したい。

(1”)について，オリジナル作文では，日本では「いつでもどこでも喫煙できる」と書かれているわけだが，この場合の「喫煙できる」は能力ではなく法律や状況的な許可に関わる概念なので，may や be allowed のほうが自然だろう。しかし，人手校閲・自動校閲とも，can を誤りとしてない。なお，日本では，現状でも，路上や公共施設では喫煙が禁じられているため，「いつでもどこでも喫煙できる」としたのは書き手の誤解であった可能性が高い。仮にそこまで汲んで修正するならば，助動詞を削り，Smokers smoke…ないし People smoke…（喫煙者はいつでもどこでもタバコを吸っている）のようにすることもできただろう。

(2”)について，「いつでもどこでも」という意味での every where every time には，表記の問題（everywhere が分綴されている）と，意味の問題（every time は「毎回」となって論旨に合わない）がある。表記の問題は人手校閲・自動校閲ともに正しく修正されている。意味の問題は，人手校閲では at any time と適切な修正がなされているが，自動校閲では誤りとされていない。なお，この部分は，anytime, anywhere などの言い方にすることもできたであろう。

(3")は, L1 干渉による時制の誤りである。日本語の条件節中では, 過去の事柄でなくても「もし～した ら」のように過去形で表現することがある。書き手はおそらく日本語からの連想で if they wanted to…としたものと思われる。人手校閲・自動校閲ともこれを誤りとして検出しているが, 人手校閲が if they want to…のように正しい時制に修正しているのに対し, 自動校閲ではこの箇所の wanted を仮定法過去ととらえ, 後続部のほうを would go to…に修正している。しかし,「食後にタバコを吸いたくなったら」という内容を仮定節で表現する必然性はなく, この修正は適切ではない。この問題については, 日本語の影響を考慮すれば正しい修正ができたと思われるが, 自動校閲は世界中のユーザーを対象にしており, 書き手の L1 への配慮はなされない。

最後に(4")について考えたい。ここには, despite の用法の問題(despite は, despite A, B の形で用いるもので, despite A の部分だけを独立して文とすることはできない)と, ロジックの問題がある。前者の問題は人手校閲・自動校閲とも誤りとして検出されており, 人手校閲は despite で始まる第 6 文を前節する第 5 文につなげることで, 自動校閲は despite を although に置き換え, 第 6 文を後接する第 7 文につなげることで, ともに, 文法的な誤りはいちおう解消されている。

しかし, この箇所のロジックの問題は解消されていない。オリジナル作文の第 5～7 文の論旨展開を確認すると,「喫煙者はいつでもどこでも喫煙できる」→「短期間なら喫煙したいという欲望を我慢できるにも拘わらず」→「煙草を吸いたいなら, 食後すぐに喫煙所に行くことができる」となる。しかし, これらの 3 つの情報は, もともと噛み合っていないため, 人手校閲のように修正しても(「喫煙者はいつでもどこでも喫煙できる」→「短期間なら我慢できるにも拘わらず, (レストランで)タバコを吸いたいなら食後に喫煙所に行ける」), 自動校閲のように修正しても(「短期間なら我慢できるが, いつでもどこでも喫煙できる」→「(レストランで)タバコを吸いたいなら食後に喫煙所に行ける」), 結局, ロジックは通らない。この箇所をロジックが通るように校閲するならば,「喫煙者は短期間なら喫煙を我慢できる」→「仮に我慢できなくても, レストランなら, 食後すぐに喫煙所に行くこともできる」→「にもかかわらず, 喫煙者はいつでもどこでも喫煙している」となるのが自然だろう。なお,「短期間」を示すために, オリジナル作文は like class(授業中？)と書いているが, そもそも授業中に喫煙することはほとんどありえないので, ここは during a meal(食事中)などとすべきであるし, 言い切り避けるにはヘッジも必要であろう。以上をふまえた校閲としては, Many of the smokers can control their desire to smoke if it is for a short time like during a meal. In addition, at restaurants, smokers can move to a smoking place soon after a meal. Nevertheless, they smoke practically anytime, anywhere.などが考えられる。

以上で, 初級学生の英作文の冒頭 7 文の校閲状況を見てきた。誤りの検出・修正の両面において, 人手校閲が自動校閲を上回ることが多かったが, そもそも, 何れの校閲であっても, オリジナルの英文を文法的に誤りがなく, 読んで自然に意味が取れるものにするという点ではきわめて不十分であった。ICNALE Edited Essays では, 専門校閲者に対して, 内容を変えずに言語的に"fully intelligible"にするよう指示を出していたため, 校閲者が踏み込んだ修正を控えた可能性はあるが, そもそも, どこまでが言語でどこからが内容なのか, 表層の文法形式は正しいものの, 読んで意味の通らない英文を "intelligible"と呼べるかどうかは議論の分かれるところである。

あえて両者を比較すれば，人手校閲が自動校閲をいくぶん上回るとしても，時間とコストをかけた人手校閲の結果がこのレベルにとどまるのであれば，むしろ自動校閲を使ったほうが合理的だという判断も出てくるだろう。今回の実例分析に限って言えば，表層的な誤りの検出・修正だけであれば，自動校閲もある程度信頼することができるが，ロジックや論旨を正しく直すという点では， 自動校閲も人手校閲も望まれる段階には達しておらず，英語教師が日常的に行っている丁寧な添削指導に代わるものにはなっていないと思われる。

6. まとめ

以上，本研究では，4つのリサーチクエスチョンに即して，2 種の校閲の差異を概観してきた。ここでは，一連の分析で得られた結果をまとめておきたい。

まず，RQ1（全体の校閲量）については，人手校閲＞自動校閲となることがわかった。また，人手校閲と自動校閲間で，40 本の作文の校閲量の相関度は r=.687（説明力約 48%）となり，同じ作文であっても，2 種の校閲の量はかなり異なることが示された。

次に，RQ2（自動校閲の観点別校閲量）については，文法（動詞活用形誤り，冠詞漏れ，名詞単複不一致など）＞句読法（コンマ抜けなど）＞スペル（複合語表記誤りなど）≒文構造（否定辞の位置誤りなど）の順になることがわかった。また，観点間の相関を見たところ，文法とスペルが，自動校閲の全体校閲量と強く相関することが確認された。

また，RQ3（書き手の習熟度との関係）については，A2 レベルの初級学生の作文は人手校閲と関連が強く，自動校閲は初級よりも中上級の作文の校閲をうまくこなす可能性が示唆された。

最後に，RQ4（校閲内容）については，人手校閲が，単数名詞の複数化・総称名詞の削除・接続詞 so の削除・動詞の多様化・ヘッジ追加等を，一方，自動校閲が単数名詞の複数化／冠詞付与／可算不可算の修正・動詞活用形修正・複合語表記の変更等をそれぞれ特徴的に行っていることが示された。

加えて，A2 レベルの初級学生の作文に対する自動校閲と人手校閲の具体的な内容を質的に検討し，それぞれの校閲は表層の文法誤りの検出と修正に関しては一定の妥当性を持っているものの，ロジックや論旨にかかる問題の修正は十分ではなく，英語教育が提供する添削指導の必要性はなお残ることを確認した。

本研究は人手校閲と自動校閲を量的・質的に比較することで，両者間には誤りの検出・修正の両面で，かなりの異なりがあることを示した。両者を比較すれば，誤りの検出・修正の点で，人手校閲がすぐれている場合が多い。もっとも，人手校閲も万能ではなく，コストを考えれば，自動校閲の使用にも一定の合理性が認められる。

自動校閲は，少なくとも現状においては，英語教師による丁寧な添削指導に取って代わるものとはなり得ていないが，工学分野では，誤りを検出するだけでなく，教師による添削に近い学習情報を盛り込んだフィードバックの自動生成の研究も進められている（永田・水本・Whittaker, 2013；永田・石川・乾, 2019）。こうした研究が展開していけば，誤り検出の精度を高めるだけでなく，より有効なフィードバックを自動で与えられる可能性も出てくるだろう。この点に関して強調しておきた

いのは，自動校閲の改善を学際的に進めていく重要性である。自動校閲システムの開発はこれまで工学者を中心に進められてきたが，工学者だけでは限界もある。水本（2016）は，工学研究における英文自動校閲の将来を展望して，「この問題を本当に解決するには，学習者がなぜ間違えるかをもっと突き詰めて考える必要がある。母語の影響，母語自体の習熟度，学習方法による影響など個々人によって異なる部分をモデル化していくと面白そうである」と述べている。この点において，多様な英語学習者の作文の実態を知悉する英語教師や英語教育学者，また，テキスト特性の抽出や分析を得意とするコーパス研究者による貢献の余地は潜在的に大きいものと思われる。今後，こうした幅広い研究者・実践者の参画を得て，自動校閲の性能がさらに上がり，学習者の英語ライティングが多面的に支援されていくことを期待したい。

注：本稿脱稿後，Grammarly ではインタフェースの大幅な変更がなされ，校閲内容にも変更が生じている可能性がある。本稿の報告内容はすべて調査時点のものである。

引用文献

浅野広樹・水本智也・乾健太郎（2017）「文法誤り訂正のためのリファレンスレス評価」『言語処理学会第 23 回年次大会発表論文集』, E6-4.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world*, *1* (pp. 91-118). Kobe, Japan: Kobe University.

Ishikawa, S. (2014). Design of the ICNALE Spoken: A new database for multi-modal contrastive interlanguage analysis. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world*, *2* (pp. 63-76). Kobe, Japan: Kobe University.

Ishikawa, S. (2018). The ICNALE edited essays; A dataset for analysis of L2 English learner essays based on a new integrative viewpoint. *English Corpus Studies*, *25*, 117-130.

Ishikawa, S. (2019). The ICNALE Spoken Dialogue: A new dataset for the study of Asian learners' performance in L2 English interviews. *English Teaching*, *74*(4), 153-177.

石川慎一郎（2012）『ベーシックコーパス言語学』東京：ひつじ書房.

久保田朗・太田学（2012）「挿入・欠落誤りを考慮した英文前置詞誤り修正支援システム」『第 4 回データ工学と情報マネジメントに関するフォーラム（DEIM2012）論文集』, F1-2

乙武北斗・荒木健治（2007）「単語出現状況の帰納的学習による英文冠詞誤りの検出及び自動校正手法」『電子情報通信学会論文誌』, *90*(6), 1592-1601.

乙武北斗・荒木健治（2010）「英文冠詞誤りの自動校正手法におけるアプローチの違いによる傾向分析」『言語処理学会第 16 回年次大会発表論文集』,PA1-38.

牧野剛典・新妻弘崇・太田学（2016）「検索エンジンによる英文の名詞語彙選択誤り検出の一手法」『第 8 回データ工学と情報マネジメントに関するフォーラム（DEIM2016）論文集』, C1-5.

松井悠太郎・新妻弘崇・太田 学(2016)「CRF による英語時制誤り検出の一手法」『第 8 回データ工学と情報マネジメントに関するフォーラム(DEIM2016)論文集』, G2-3.

松崎幸太郎・田中久美子(2006)「英文とその校正からの書き換えルール自動抽出」『言語処理学会第 12 回年次大会発表論文集』, S1-1.

水本智也(2016)「自然言語処理技術の現状と展望(エラー分析プロジェクトを通して):3.16 英文校正」『情報処理』,*57*(1),40-41.

永田亮・石川慎一郎・乾健太郎(2019)「解説文生成研究のためのライティング技術解説付き学習者コーパス」『言語処理学会第 25 回年次大会発表論文集』, P1-1.

永田亮・水本智也・E. Whittaker(2013)「前置詞誤り検出／訂正のための誤り核フレームの生成」『言語処理学会第 19 回年次大会発表論文集』, P3-12.

谷本太郁由・太田学(2011)「検索エンジンを用いた英文コロケーション誤りの検出と修正」『第 3 回データ工学と情報マネジメントに関するフォーラム(DEIM2011)論文集』, A8-4.

若菜崇宏・永田亮・桝井文人・河合敦夫(2005)「可算／不可算判定を用いた英文の冠詞誤り検出」『言語処理学会第 11 回年次大会発表論文集』, P5-15.

Wordvice (2016) 「英文校正ワードバイスライティングレポート 2016」Retrieved from wordvice.jp/2016-レポート-英文校正-データ