

変更履歴を用いた英語文章作成支援システムの実現

Implementing a Support System using a Revision History for English Composition

徳岡 拓弥*¹ 竹内 章*² 國近 秀信*²
Takuya Tokuoka Akira Takeuchi Hidenobu Kunichika

*¹ 九州工業大学大学院情報工学府
Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology

*² 九州工業大学大学院情報工学研究院
Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology

When novice learners of English compose English sentences, experts revise the sentences to correct errors. Learners, however, sometimes make the same errors repeatedly. In order to solve the problem, we propose a method to detect errors by using revision history, so that the learners can correct errors themselves. In this paper, we describe the implementation of such a support system that records and retrieves the revision history.

1. はじめに

英語の初学者が英語文章を書く場合、熟達者による添削が繰り返し行われる場合が多い。その際には、何度も同じ誤りを繰り返す場合があり、添削者の負担となると考えられる。これまでに多くの支援システムが実現されているが(例えば、[大野2008])、それらの多くはあらかじめ用意した誤りに関する知識を用いて誤りを同定・指摘する。それに対し本研究では、文章作成時に過去の添削履歴を提示することより学習者自身に変更を促すというアプローチをとる。本研究では、Microsoft社のWord 2007以降(以下、Word)の変更履歴機能を利用した添削を前提とし、履歴の抽出および検索をおこなう英語文章作成支援システムの実現を目的とする。

2. 英語文章作成支援

誤りを繰り返す主な原因としては、英語の文法や語法を理解していない、指摘されたことは覚えているがたまたま間違えたなどが考えられる。本研究では、過去の添削履歴の確認と学習者自身による修正とを繰り返すことにより、その後の誤りの繰り返しが減少すると考え、学習者が英語文章を記述する際に過去の添削箇所と同じ表現を用いた場合にその変更履歴を提示することで支援をおこなう。その際には、英文法/語法の理解促進のため、単に添削箇所および内容を提示するだけではなく、添削の要因となった語句も合わせて明示することが望ましい。なお本研究で対象とする添削者による変更は、削除、挿入、コメントおよび修正であり、それぞれWordの変更履歴における削除、挿入、コメント、同一箇所の削除および挿入に対応する。

3. 英語文章作成支援システムの実現

図1に本システムの概要を示す。本システムは、変更履歴抽出機能、変更履歴データベースおよび変更履歴検索機能から成る。まず学習者はWord上で英語文章を作成した後、添削者へファイルを提出する。次に添削者は、Wordの変更履歴機能を用いて文章を添削し、学習者へ返却する。続いて、学習者が変更履歴抽出機能を用いて変更履歴をデータベースへ蓄積する。その後、学習者が英語文章を書く際に変更履歴検索機能を用いて

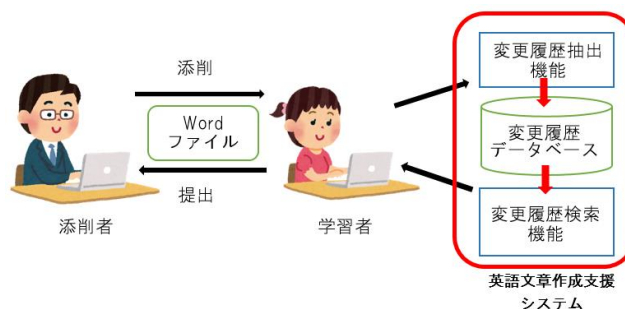


図1 システムの全体図

利用することにより、作成中の英語文章をもとに本システムが変更履歴を検索し、ヒットした場合にその情報を学習者へ提示する。以下、各構成要素について述べる。

3.1 変更履歴抽出機能

本機能は、添削結果に相当する変更履歴およびその関連情報を抽出する。過去に作成した誤文と同一の文を作成するとは限らないため、誤りに関する情報を一般化し、蓄積する情報の適用範囲を拡大する必要がある。そのための一つの方法として、変更箇所およびその要因となった語句のみを蓄積することが考えられるが、現時点では変更の要因となった語句ではなく、変更箇所と係り受け関係にある語句のヘッドを関係語として利用する。

具体的にはまず、Wordのdocx形式のファイルよりdocument.xmlおよびcomment.xmlを取り出し、それらの中から特定のXMLタグを手掛かりとして変更前後の語句を文単位で抽出する。次にStanford parser [Klein 2003]を用いて抽出した英文を構文解析し、語句の品詞および文の構造を同定する。その後、構文解析結果より、変更箇所、関係語、および、最も近い共通の祖先の品詞を抽出する。関係語の抽出方法[福島2013]は、以下の通りである。

- 変更箇所がヘッドではない場合、当該単語を含む句のヘッドを関係語として抽出する。

連絡先: 徳岡拓弥, 九州工業大学大学院情報工学府,
〒820-8502 福岡県飯塚市川津 680-4,
tokuoka@minnie.ai.kyutech.ac.jp

- 変更箇所がヘッドの場合、その兄弟語句のヘッドを関係語として抽出するとともに、親ノードについて再帰的に関係語の抽出処理をおこなう。ただし親ノードがヘッドではない場合は、当該ノードの処理をおこなった後に終了する。

例として、“I made some big box.”が“I made some big boxes.”へ変更された場合を考える。変更前の英文の構文木を図2に示す。変更箇所であるboxはNPのヘッドであるため、その兄弟のsomeおよびbigが関係語として抽出される。また、親ノードが再帰的に処理され、VPのヘッドではないためVPのヘッドであるmadeが関係語として抽出される。よって、最終的に抽出される情報は、(NP (DT some) (NN box)), (NP (JJ big) (NNS box))および(VP (VBD made) (NN box))となる。

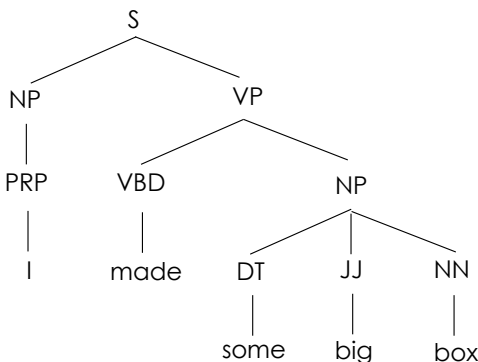


図2 “I made some big box.”の構文木

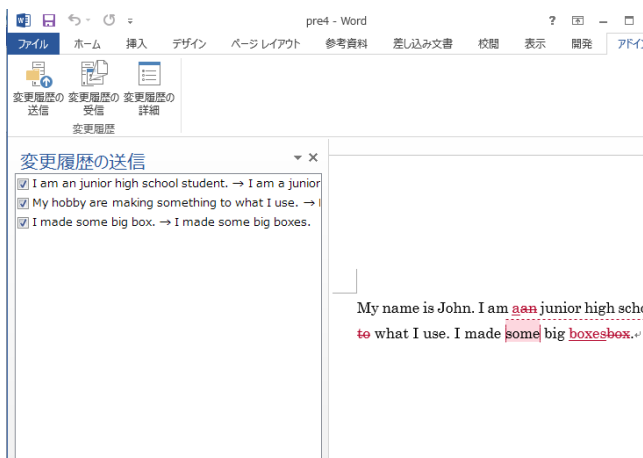


図3 変更履歴抽出機能の画面例

図3に変更履歴抽出機能の画面例を示す。右側のウィンドウは通常のWordの画面であり、左側のウィンドウには抽出された変更箇所を含む英文が表示される。学習者が「送信」ボタンを押すと、次節で述べる変更履歴データベースに蓄積される。

3.2 変更履歴データベース

本データベースは、添削結果に相当する変更履歴の蓄積、および、英語文章作成時における過去の変更履歴の検索のために利用される。検索時には前節で述べた単語とその関連語を含む構文木が検索キーとして利用され、ヒットした場合には変更前後の英文を表示する。そのため、本データベースでは、変更箇所とその関係語を含む構文木、変更前後の英文、変更の種類、およびコメントを保持する。

3.3 英語文章作成支援機能

誤りの繰り返しを防ぐため、学習者が英語文章を作成する際に過去の添削履歴を参照する機能である。処理の流れは次の通りである。まず document.xml より文単位で英語文章を抽出して構文解析し、文章中の各単語およびその関係語を含む構文木を生成する。次にその構文木を検索キーとして変更履歴データベースを検索し、ヒットした場合にはその結果を学習者へ提示する。

図4に実行結果の例を示す。中央のウィンドウは学習者が英語文章を作成する画面であり、過去の変更履歴と一致した場合、左下のウィンドウに単語とその関係語を含む構文木のリストが表示される。学習者が一つの構文木を選択すると、左上のウィンドウへ変更前後の英文が表示される。さらにその中の一つを選択すると、右側のウィンドウへ変更前後の英文、変更の種類、コメント等が表示される。

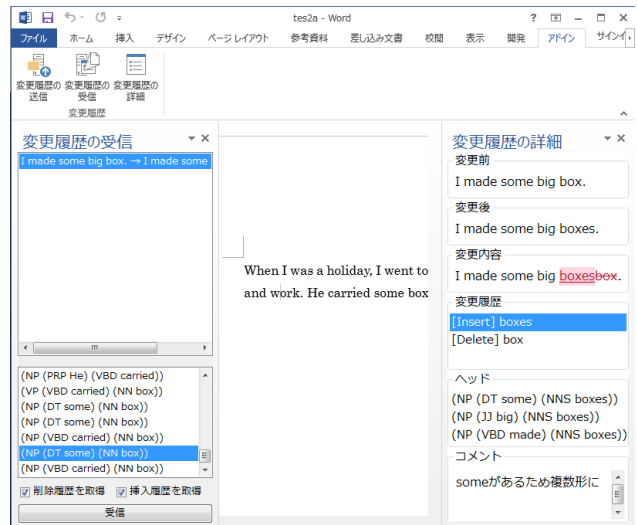


図4 変更履歴検索機能の画面例

4. おわりに

本論文では、変更履歴を用いた英語文章作成支援システムの実現について述べた。本システムを用いて、特定の添削者または特定分野の複数の添削者による添削履歴を蓄積し利用することにより、人や分野に依存した英語文章作成支援が可能になると考える。

現時点では、検索結果を単純に表示しているため、多くの変更履歴がヒットした場合には必要な情報へ辿り着くことが困難になると考えられる。従って、今後は検索結果の集計等も含めた変更履歴の参照支援機能の実現が必要である。さらに学習者の負担軽減のため、変更履歴の蓄積および検索の自動化を予定している。

参考文献

[福島 2013] 福島良典: 文章作成支援システムにおける変更箇所抽出部の改良, 2012 年度九州工業大学卒業論文, 2013.

[Klein 2003] Klein, D., Manning, C.: Fast Exact Inference with a Factored Model for Natural Language Parsing, *Advances in Neural Information Processing Systems 15*, pp.3-10, 2003.

[大野 2008] 大野博之, 稲積宏誠: 技術文章の校正・推敲支援ツールにおける機能拡張容易性の向上, 信学技報 ET2007-89, pp. 31-36, 2008.