Vol.2014-DBS-159 No.2
Vol.2014-IFAT-115 No.2
2014/8/1

Wikipediaにおける単語の順序を考慮した編集の差し戻し検知手法

近藤 弘隆[1,a)]　中村 晃[1,b)]　鈴木 優[1,c)]　石川 佳治[1,2,d)]

概要：本稿では，記事中での単語の順序を考慮し，Wikipedia における編集の差し戻しを検知する方法を提案する．既存研究では，編集された単語を単語集合として記録し，単語集合の一致を差し戻しの基準としている．しかし，この手法では，単語集合は一致するが，編集箇所の異なる場合に対して，差し戻しを誤検知してしまう可能性がある．本研究では，単語の位置を考慮して，編集の差し戻しを検知する．単語の位置を考慮するにあたり，追加された単語は記事上に存在する単語であるため，削除されるまで記事中での位置を一意に定めることができる．一方で，削除された単語の位置は，記事上には存在しないため一意には定まらない．本手法では，単語が削除された版から再び追加された版までの差分情報を用いて，削除された単語の推定位置を算出している．評価実験を行った結果，既存手法に対して精度の向上を確認した．

キーワード：Wikipedia，差し戻し検知，編集履歴，編集行動

1. はじめに

Wikipedia は Web 上の百科事典である．誰でも編集できるという特徴を持っている．Wikipedia では，過去の編集履歴データが全て公開されている [*1]．これは，Wikipedia の全ての記事に対して，記事が作成された版から現在の版までの各版における記述と，その編集者を含むデータである．この編集履歴を解析し，抽出されたデータを利用した研究が行われている．例えば，編集履歴から編集者の行動データを抽出することにより，編集者の質や記事の質を推定するといった研究が存在する．我々は，この「編集履歴から編集者の行動を抽出する」という部分に着眼した．

編集者の行動とは，記述の追加と削除のことである．また，本稿では便宜上，記述を単語単位で扱う．すなわち追加行動は，「編集者 e が単語 w を追加した」と定義する．削除行動は，「編集者 e_d が，編集者 e_i によって追加された単語 w を削除した」と定義する．追加行動と削除行動を合わせて**編集行動**と呼ぶ．

前述のような，編集者と記事の質推定などの編集行動データを利用した研究は，編集履歴からのデータ抽出が正確に行われていることを前提としている．したがって，提案された手法の精度は，編集行動データの正確性に依存しているといえる．しかし，単純な手法では編集行動データの抽出を正確に行うことはできない．単純な手法とは，前述のように，記事の最初の版から順に各版の差分を取得することによって，編集行動を抽出する手法のことである．このような手法を用いても正確な編集行動を抽出できない原因は，**差し戻し**とよばれる編集が存在するためである．

差し戻しとは，以前に行われた一つ以上の編集を完全に取り消す行為を含む編集のことである．ある編集 E_s が別のある編集 E_t を差し戻すとは，E_t において追加された単語が全て E_s において削除され，かつ E_t において削除された単語が全て E_s において追加されることである．このとき，E_s を**差し戻し元**，E_t を**差し戻し先**とよぶ．また，E_s と E_t の編集ペアを**差し戻しペア**とよぶ．

差し戻しには，記述の追加行動を取り消す（削除する）場合と，記述の削除行動を取り消す（再び追加する）場合がある．後者の際には，その記述を書いた人が誰かという情報が，もとの追加者から，削除後の二度目の追加者にすり替わってしまうという問題がある．したがって，単純な手法においては，差し戻しが行われた場合に正確な編集行動データを抽出できない可能性が生じる．編集履歴から，より正確な編集行動を抽出するためには，差し戻しを検出し，その影響を考慮する必要がある．より高精度に差し戻

[1] 名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University
[2] 国立情報学研究所
National Institute on Informatics
[a)] kondou@db.ss.is.nagoya-u.ac.jp
[b)] nakamura@db.ss.is.nagoya-u.ac.jp
[c)] suzuki@db.ss.is.nagoya-u.ac.jp
[d)] ishikawa@is.nagoya-u.ac.jp
[*1] http://dumps.wikimedia.org/

表 1 版ごとの記述

版	記事本文
1	オレンジ
2	オレンジ メロン バナナ
3	オレンジ メロン バナナ グレープ ピーチ
4	オレンジ グレープ ピーチ
5	オレンジ

表 2 編集ごとの編集された単語

編集	編集前の版 / 後の版	追加された単語	削除された単語
E_a	0/1	オレンジ	
E_b	1/2	メロン バナナ	
E_c	2/3	グレープ ピーチ	
E_d	3/4		メロン バナナ
E_e	4/5	オレンジ	グレープ ピーチ

しを検出することが，より正確な編集行動の抽出につながり，前述のような編集履歴データを利用した研究の精度向上に貢献すると考えられる．

差し戻しは特定の編集の取り消しを行い，記述を編集前の状態に戻す．既存研究では，記述全体が過去の版の記述に戻っているかを基準として差し戻しを検出する手法が提案されている．この方法で検出できるのは，二つの版における記述が完全に一致する場合だけである．しかし，記述が完全に一致せずとも，実際には差し戻しを行っている場合が存在する．差し戻し元と差し戻し先との間に，何らかの編集が行われていた場合，差し戻し元の編集後の記述と，差し戻し先の編集前の記述が，完全には一致しないため，差し戻しの検知漏れが生じる．この方法によって検知された差し戻し先と，差し戻し元との間に，**編集箇所の重なる編集ペア**が存在した場合，差し戻しの誤検知が発生してしまう．編集箇所の重なる編集ペアとは，ある記述を追加する編集とその記述の一部の削除する編集のペア，もしくは，ある記述を削除する編集とその記述の一部を再び追加する編集のペアのことである．この方法による差し戻しの検知漏れと誤検知の例を**表 1** と**表 2** を用いて示す．表 1 は，ある記事における版ごとの記述を表している．表 2 は，編集ごとに追加された単語と削除された単語を示している．版 1 と版 5 の記事本文が完全に一致しているため， E_b, E_c, E_d が，編集 E_e によって差し戻されたと検知される．E_b の「メロン」と「バナナ」を追加する編集は，直感的には E_d により差し戻されていると判断できる．しかし，E_c で「グレープ」と「ピーチ」の追加があるため，版 1 と版 4 が一致せず，この手法では E_b と E_d のペアは差し戻しとは検知されない．また，版 1 と版 5 のように記述全体が一致した版のペア間に，E_b と E_c のように編集箇所の重なる編集ペアが存在すると，差し戻しを誤検知してしまう．E_b と E_c は同じ単語である「メロン」と「バナナ」に対する編集であり，E_b ではこれらを追加し， E_c ではこれらを削除している．E_e ではこれらの単語の追加も削除もしていない．つまり， E_e は E_b や E_c に対して一切の差し戻しを行っていない．しかし，この手法では， E_e は E_b と E_c を差し戻したと検知してしまう．このように記述全体の一致を差し戻し検知の基準とすると，差し戻しの検知漏れや誤検知が生じる可能性がある．

上述のような検知漏れや誤検知を防ぐために，記述全体ではなく，記述の一部を差し戻しの判定基準とする必要があると考えられる．そこで本研究では，記事の最小構成単位である単語に着目し，差し戻しの判定基準を単語単位とした．本研究と同様に，単語に着目した既存研究が存在する．その研究では，追加・削除された単語の表層文字列と各単語が追加・削除された数をもとに判定を行っている．編集された単語を追加された単語集合 W_a と削除された単語集合 W_r の組で表現したとする．この手法では，ある編集における W_a と W_r が，それ以前の編集における W_a と W_r を含むとき，差し戻しと判定する．しかし，単語単位で差し戻しの判定を行うときに，単語をその表層文字列だけで識別した場合，編集の単語集合が偶然一致したときに，誤検知する可能性が生じる．たとえば，ある単語を削除した後に，別の版で同じ表層文字列の単語が追加されたとき，その削除行為が差し戻しであるとは限らない．なぜならば，単語の表層文字列が一致していたとしても，単語そのものは一致しておらず，同じ表層文字列の単語を別の文に追加しただけといった場合が想定されるからである．本研究では，単語を一意に定めるため，単語の表層文字列だけでなく，記述中での位置を利用している．過去に追加された単語と同じ位置，同じ表層文字列の単語が削除されれば，追加行動に対する差し戻しとする．また，過去に削除された単語の，現在の版における推定位置に，同じ表層文字列の単語が追加された場合は，削除行動に対する差し戻しとする．本稿では，このように単語の位置を考慮した編集差し戻しの検出手法を提案する．

2. 関連研究

従来より，様々な目的に応じた差し戻しの検知が行われている．例えば，Kittur ら [1] や Halfaker ら [2] は 編集後の版の記述全体が以前の版の記述全体と一致するかどうかを調べることよって，差し戻しを検知している．記述全体が一致した場合，その編集は一致した編集までに行われた編集全てに対し，差し戻しを行っているということになる．

Flöck ら [3] は，Kittur らの手法で検知される差し戻しは，実際に Wikipedia の編集者が考える差し戻しとは一致しない場合があると指摘し，それに代わる手法 (DIFF) を提案している．Kittur ら [1] や Halfaker ら [2] の手法は，記述全体が過去の版の記述に戻っているかどうかを調べる方法であるため，1 章で述べた欠点が存在する．DIFF では記事本文の一致を見るのではなく，編集で追加，削除された単語に着目している．記事が作成された時の版を第 0 版と

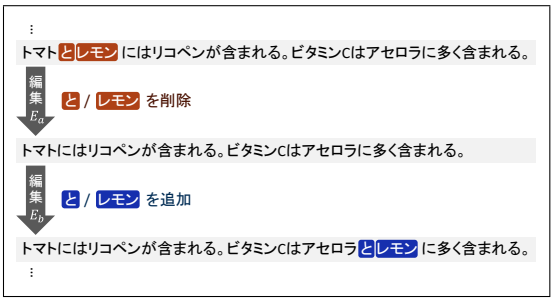

図 1 差し戻しではない編集の例

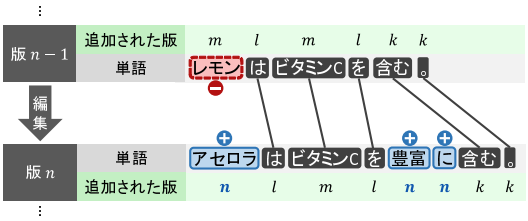

図 2 差分イメージ図

する．まず s 番目の版と $s-1$ 番目の版との差分を取り，編集 E_s で追加された単語，削除された単語を Bag-of-Words として保持する．次に， E_s 以前の編集に対し，新しい順に，次の手順で差し戻されたかどうか判定していく．ある編集 $E_t(0 \leq s-i \leq t \leq s-1)$（$i$ は 0 以上の整数）に対し，E_t において追加された単語集合 W_a^t が E_s において削除された単語集合 W_r^s の部分集合であり， かつ E_t において削除された単語集合 W_r^t が E_s において追加された単語集合 W_a^s の部分集合である時， E_s は E_t を差し戻したとする．そして，差し戻しであったとき，$W_r^s := W_r^s \setminus W_a^t$, $W_a^s := W_a^s \setminus W_s^t$ とする．そしてさらに以前の編集と比較していく．このように差し戻しを検知していく．

Bag-of-Words は単語の種類と数だけを考慮し，記事中での単語の位置は考慮していない．そのため，1 章での，単語をその表層文字列だけで識別した場合と同様の欠点を持つ．**図 1** では，DIFF で誤検知してしまう編集のペアの例を示している．図 1 では編集 A は「と」と「レモン」を削除している．編集 B は「と」と「レモン」を追加している．DIFF では B は A を差し戻していると判定してしまうが，実際には編集箇所が違うため，差し戻しではない．

また，Adler ら [4], [5] は記事の質を算出するために，編集距離を用いた差し戻し検知手法を提案している．この編集距離を用いる手法は，最終的に編集が取り消されたことを正確に検知することは可能であるが，ある編集の全てが一度の編集で取り消されたかどうかの検知に使用する際には使用できない．

3. 提案手法

前章で DIFF には差し戻しを検知する際に誤検知の可能性があることを述べた．この誤検知はある編集で追加された単語と削除された単語を記録する際に，単語の位置を考慮していないことによって発生する．ここでの単語の位置とは，記事を単語の列と考えた時の先頭からのその単語の順序とする．

ある編集 E_a が以前の編集を差し戻しているかどうかを判定する際， E_a によって追加された単語が，以前削除された単語に対応する単語であるのか，新たに追加された単語であるのか判断することは困難である．これは削除された単語は削除された時点で記事上から存在しなくなるためである．本手法では追加された単語が削除された単語と対応すると対応するかどうかを判定するために，削除された単語を**推定位置**（削除されなかったら存在したであろう位置）を用いて保持している．

位置を考慮した比較を行う場合，記録している編集で追加された単語の位置，削除された単語の推定位置を E_a が行われた後の版でどの位置にあたるか更新する必要がある．そこで，本手法では，編集前後の二つの本文の差分を取ることによって，どの位置の単語が残り，追加され，削除されたのかを求め，単語の位置を更新していく．

編集履歴を古いものから順に，それ以前の編集を差し戻しているか判定していくことによって，差し戻しを検知する．

3.1 節では編集の記録方法を述べ，3.2 節ではその記録を用いた差し戻しの検知手法について述べる．

3.1 編集の記録方法

ある編集を記録するために，編集で追加された単語とその位置，削除された単語とその位置を取得したい．追加された単語に関しては，記事本文に存在する単語であるため，後の版でどこに存在するのかという対応関係を算出することが可能である．しかし，削除された単語は削除され，記事本文には存在しないため，対応関係がわからない．そこで削除された単語の位置を推定位置で表現する．

Wikipedia の記事の編集履歴からは，各版での記事本文を得ることができる．まず，編集の行われた後の版と編集の行われる前の版の記事本文の差分を取る．記事本文の差分を取ることによって，編集後の版のどの位置に単語が追加され，編集前の版のどの位置にあった単語が削除され，どの位置の単語が編集されなかったのかを算出する．**図 2** は，差分をどのように算出しているのかという図である．図 2 はある編集の編集前の文章と編集後の文章とその対応関係を示すものである．図 2 の編集では，前後の版に存在する「は」，「ビタミンC」，「を」，「含む」，「。」を基準に，編集前の版だけに存在する「レモン」を削除し，編集後の版だけに存在する「アセロラ」，「豊富」，「に」を追加したという情報を算出している．

次に，編集で追加された単語の情報として，追加された

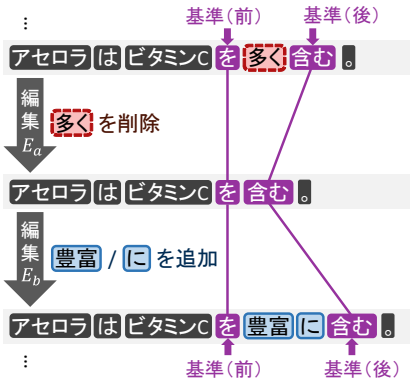

図 3 削除された単語とその推定位置の基準

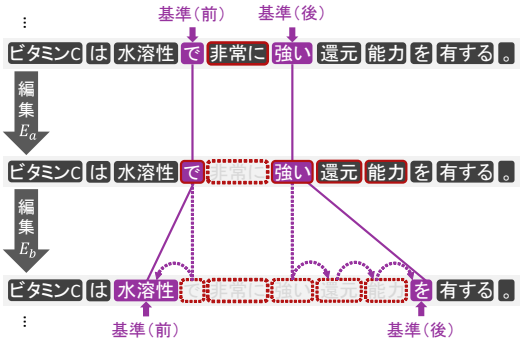

図 4 削除推定位置の基準更新

単語とその位置，追加された版を記録する．また，削除された単語の情報として，表層文字列と削除された版，その単語の推定位置を記録する．この削除された単語の推定位置は二つの単語に挟まれる範囲として表現する．その二つの単語とは，編集前から残っている，削除された単語の前後に存在する一番近い単語である．**図 3** は削除された単語の基準となる単語の変遷を示した図である．図 3 では削除された単語「多く」の前後に存在する変化しなかった単語である「を」と「含む」を基準の単語としている．

また，この版までに追加された単語の情報を用いて，編集後の記事本文に存在する単語それぞれについて，どの版で追加された単語であるのかを算出し，今までに追加された単語の位置を更新する．今までに削除された単語の推定位置の更新も行う．

3.1.1 節では今までに追加された単語の位置を更新する手法について，3.1.2 節では今までに削除された単語の推定位置を更新する手法について述べる．

3.1.1 追加された単語位置の更新

今までに追加された単語位置の更新は，記事に存在する単語がどの版で追加されたのかを求めることによって算出することが行なえる．追加された単語のうち，記事に存在しない単語はすでに以前の版で削除され差し戻されているからである．

差分を用いて，編集後の記事中のどの単語がどの版で書かれたのかを以下のように求める．編集前の版での単語がどの版で書かれた単語かといった情報があるとする．編集前後の記事の差分を取ることにより得ることができる情報と，編集前の版での単語がどの版で書かれた単語かといった情報を用いて，編集後の記事中に存在する単語がどの版で書かれたのかを算出する．図 2 において編集後において追加された単語「アセロラ」，「豊富」，「に」は編集後の版で追加した単語である．変化しなかった単語「は」，「ビタミンC」，「を」，「含む」，「。」に関しては，編集前の情報と照らし合わせることによりどの版で追加された単語なのかといった情報を得ることができる．図 2 では，これらの差分から得られた情報を利用して編集後の記事中のどの単語がどの版で書かれたのかを算出している．また，記事が作成された版（最初の版）に関しては，編集前の版は存在しない．最初の版に存在する単語は，全て最初の版で追加されたと考える．

このように，ひとつ前の版における単語がどの版で書かれたのかといった情報を用いることにより，どの位置の単語がどの版で書かれたのかという情報を再帰的に求めることができる．この情報をもとに，今までに追加された単語の位置を更新する．

3.1.2 削除された単語位置の更新

削除された単語の推定位置を更新する手法について述べる．削除された単語の推定位置は，記事に存在する二つの単語の位置に挟まれる範囲として定義した．編集が行われると，その削除された単語の位置の基準となる単語の位置が移動することがある．移動後の位置を，差分を取ることによって求めた編集前後の単語の対応関係を用いて，編集後の基準となる二つの単語の位置を更新する．図 3 では編集 B で「豊富」と「に」が追加されたことを受けて，後ろの基準となる「含む」の位置が二つ後ろへ移動している．また，基準となる単語が削除された場合，削除されたのが前側の基準となる単語であった場合は基準となっていた単語より文頭に近い単語の中で最も近い変化しなかった単語が，新たな基準の位置になる．一方，後ろ側の基準となる単語が削除された場合は削除された単語より文末の単語の中で一番近い変化しなかった単語が新たな基準の位置になる．**図 4** は編集によって削除された単語の推定位置を表す基準となる単語が削除された場合の図である．図 4 では前側の基準となっていた「で」が削除されたため，それより文等に近い単語で変化しなかった単語の内，一番近い「水溶性」が新たな前側の基準となっている．同様に，後ろ側の基準であった「強い」が削除されたため，それより文末に近い単語で変化しなかった単語の内，最も近い「を」が新たな前側の基準となっている．

3.2 差し戻し検知手法

ある編集によって追加された単語が以前に削除された単語を差し戻している場合，その追加した単語は以前に削除された単語の表層文字列が同じであり，追加された位置は削除された単語における推定位置の範囲内であるはずである． また，ある編集によって削除された単語は，その削除された単語を追加された編集について差し戻している．

編集 E_a が行った全ての追加，削除に対し，編集 E_b 上記の条件を満たす場合， E_b は E_a を差し戻していたと検知する．

節 3.2.1 において削除された単語がどの版で追加された単語なのかの判定法を，節 3.2.2 では追加した単語が以前削除した単語を差し戻すものか，差し戻された単語はいつ削除されたものだったのかの判定法を述べる．この判定結果から，以前の編集を差し戻していないかを確かめる．

3.2.1 削除された単語に対する判定

削除された単語がどの版で追加された単語かの判定は以下のように行う．削除された単語というのは編集前まで記事に存在していた単語であるため，編集前のどの位置の単語が削除されたかを差分によって取得することができる．編集前に記事に存在している単語については，3.1.1 節で述べたようにどの位置の単語がどの版で書いたものかという情報を記録している．この情報を用いて，削除された単語がどの版で追加されたものかを求める．

3.2.2 追加された単語に対する判定

追加された単語が以前削除された単語だったかどうかの判定は以下のように行う．削除された単語の位置は，前述のように二つの単語に挟まれる位置として記録している．追加された単語が以前削除された単語と同じ単語であり，その単語の推定位置の基準となる二つの単語に挟まれる位置への追加であった場合に，追加した単語は以前削除した単語を差し戻しているということになる．**図 5** は削除された単語が基準位置に挟まれる位置へ再び追加される様子を表す図である． 図 5 では，追加した「アセロラ」が以前削除された「アセロラ」の基準の単語「は」と「に」に挟まれる位置への追加であるため，削除した単語を差し戻していると判定される．差し戻した単語の削除した版の情報を取得することによって，どの版への差し戻しであるかを算出する．

4. 評価実験

4.1 実験概要

日本語版 Wikipedia の編集履歴データ [*2] を利用して，提案手法と DIFF[3] の比較実験を行った．両手法でともに検知された差し戻しに対して，人手による評価を行った．なお，差し戻しを検知する際に保持する編集を 20 件とした．

[*2] http://dumps.wikimedia.org/jawiki/

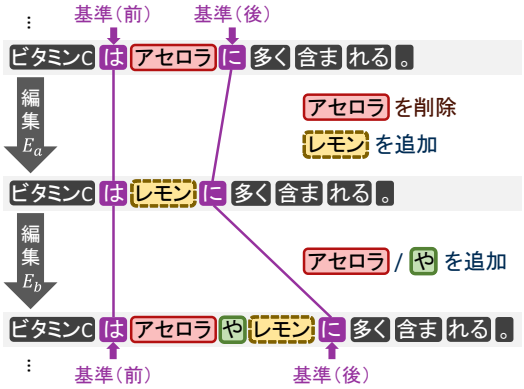

図 5 削除推定位置への追加

表 3 差し戻し数

		DIFF	
		検知した	検知していない
提案手法	検知した	79	3
	検知していない	27	不明

4.2 実験内容

4.2.1 実験データ

実験に用いる編集履歴のデータとして，2014 年 4 月 16 日の Wikipedia 日本語版 Dump データ [*3] から記事 「エジプト」の編集履歴を用いた．編集履歴から提案手法と DIFF それぞれを用いて差し戻しを検知した．4 月 16 日時点での Dump データに記録されている記事「エジプト」の編集履歴には 778 件の版が存在する．778 件の版から考えられる編集の組み合わせは 30 万組を超えるため，全ての組に対して実際の差し戻しであるかを人手によって判定することは困難である．そこで両手法で検知することのできた差し戻しに対してだけ，実際の差し戻しであるかを人手によって評価した．評価の際には，5 人の作業者に判定を行わせ，全員の多数決を取った．

4.2.2 評価指標

評価指標として，適合率（検知された差し戻し数に対する，正解数の割合），再現率（全正解数に対する，検知された正解数の割合）を用いた．前述のように，編集の組すべてに対して実際の差し戻しであるかを人手によって判定することは困難であるため，両手法によって検知された差し戻しのうち，実際の差し戻しと判定された差し戻しを全正解数とした．つまり，**図 6** の各円は，それぞれの手法で検知された差し戻しおよび実際の差し戻しの集合を表しており，実際の差し戻しの円と，他の円の重複部分が正解データ集合となる．ここでは便宜的に再現率という用語を用いているが，一般に言われる再現率とは意味が異なる．

4.3 実験結果と考察

両手法によって検知された差し戻し数を**表 3** に示す．

[*3] http://dumps.wikimedia.org/jawiki/20140416/

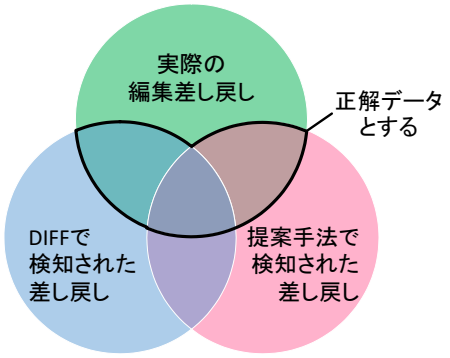

図 6 実際の差し戻しと検知された差し戻し

表 4 実験結果

	適合率（正解数/差し戻し数）	再現率（正解数/全正解数）
DIFF	70.8% (75/106)	100% (75/75)
提案手法	90.2% (74/82)	98.7% (74/75)

実験結果を**表 4** に示す．表 4 より，提案手法の適合率は DIFF と比べて高い結果となった．また，DIFF で検知され，提案手法で検知されなかった差し戻しのうち，実際に差し戻しであったものは 1 件存在した．取得できなかった差し戻しに関しては，編集前後における記事本文の単語の対応関係を算出する際，単純な編集距離が最小になる差分を求めるのではなく，文脈を考慮する手法を用いることにより改善できるのではないかと考えられる．

また， DIFF によって検知された差し戻しのうち，提案手法によって検知されなかった差し戻しに関しては，98.7%が差し戻しではないという結果になった．

DIFF によってだけで検知された差し戻し先のうち，実際には差し戻されていないと判定された編集は，編集単語数が少なく，また，編集された単語は頻繁に編集される単語であった．これは，編集した単語が偶然後の編集と一致してしまったためであると考えられる．例えば，**図 7** で編集されている単語は「[[」と「]]」であり，これは Wikipedia 内の記事へリンクを張る際に使用されるものである．この単語は頻繁に追加，削除されるため，DIFF では誤検知されてしまっている．このような誤検知は位置を考慮することによって，違う位置への追加，削除を区別することができるため，提案手法では誤検知を減らすことができた．

提案手法による誤検知が発生する原因としては，削除された単語の推定位置の範囲が編集を繰り返すことにより広がることが考えられる．そのため，文脈的には削除された単語と関係なく追加された単語が，削除された単語の推定位置と偶然一致してしまい，誤検知してしまったと考えられる．推定位置が広がる原因には，削除された単語の位置を表す基準となる単語が削除された場合や，削除された単語の推定位置の範囲内に単語が追加されるということが考えられる．

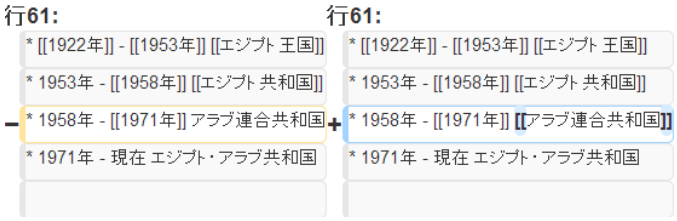

図 7 編集単語数の少ない編集

5. まとめ

本稿では，Wikipedia における正確な編集行動の抽出のために，編集履歴データから差し戻しを検知する手法を提案した．提案手法では差し戻しかどうかを厳密に判定するために，追加された単語の位置，削除された単語の推定位置を求め，編集を記録している．評価実験の結果，提案手法の適合率は既存手法より高い結果となった．

本手法は単語ごとに位置を管理しているため，本手法をより発展させることにより，単語一つ一つが差し戻されたかどうかという判定にも応用できると考えられる．単語単位で差し戻しを検知することによって，編集行動のより正確な抽出につながると考えられる．

差し戻し検知手法や, 編集行動を求める手法は, Wikipedia をはじめとする Wiki 形式の編集履歴だけでなく，バージョン管理システムに応用することが考えられる．複数人が同時に使うバージョン管理システムに本手法を用いて，誰がどの部分を書いたのかを求めるシステムを作成することができる．

謝辞 本研究は JSPS 科研費 25280039, 23700113 の助成を受けたものです．

参考文献

[1] Kittur, A., Suh, B., Pendleton, B. A. and Chi, E. H.: He Says, She Says: Conflict and Coordination in Wikipedia, *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, pp. 453–462 (2007).

[2] Halfaker, A., Kittur, A., Kraut, R. and Riedl, J.: A Jury of Your Peers: Quality, Experience and Ownership in Wikipedia, *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, WikiSym '09, ACM, pp. 15:1–15:10 (2009).

[3] Flöck, F., Vrandečić, D. and Simperl, E.: Revisiting Reverts: Accurate Revert Detection in Wikipedia, *Proceedings of the 23rd ACM conference on Hypertext and social media*, ACM, pp. 3–12 (2012).

[4] Adler, B. T. and De Alfaro, L.: A Content-Driven Reputation System for the Wikipedia, *Proceedings of the 16th international conference on World Wide Web*, ACM, pp. 261–270 (2007).

[5] Adler, B. T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I. and Raman, V.: Assigning Trust to Wikipedia Content, *Proceedings of the 4th International Symposium on Wikis*, WikiSym '08, ACM, pp. 26:1–26:12 (2008).