

# Project Next 日本語校正タスク

山本和英（長岡技術科学大学）

yamamoto@jnlp.org

鄭育昌（富士通研究所）

cheng.yuchang@jp.fujitsu.com

## 概要

**Project Next** 日本語校正タスクの最終報告として、日本語校正技術について議論する。まず、日本語校正タスクの特徴について議論した後、現在のシステム開発状況、及び研究状況を概観する。次に、非日本語話者の書いた日本語文章を対象テキストにして、様々な誤り事例に対して現状の自然言語処理技術で対応可能かどうかについてコメントする。5節では実際の校正システムの出力結果に対して考察を行い、6節で総括する。

## 1. 日本語校正タスクの特徴

日本語校正タスクは任意の日本語文章を入力して、誤りや不自然な部分（以後は不自然さも含めて単に「誤り」と呼ぶ）を指摘、もしくは訂正するタスクである。このタスクは、例えば機械翻訳や形態素解析のように明確に出力が定義されているタスクとは異なり、誤り検出のみの場合と訂正候補を提示する場合に分かれ、さらに候補提示の場合は単一候補と複数候補の提示に分かれる。

日本語文章の誤りは非常に多岐にわたることから、日本語校正タスクで対象とすべき項目も同様に多岐にわたる。まず手書きの場合は仮名や漢字の表記も（本来であれば）校正対象に含まれる。電子化文書に限定しても、表記（用言の活用や送り仮名、ローマ字表記等）、語彙の選択、助詞、文法、意味的な整合性や文体、談話構造など、日本語に関するありとあらゆる項目が誤りの対象（＝校正の対象）となり得る。

これに関連して、他の自然言語処理タスクと大きく異なるのは、処理の前提で形態素解析ができない可能性があるという点である。文字認識や音声認識、あるいは Web 上のネットスラング解析などごく一部のタスクを除き、自然言語処理タスクのほとんどは「正しい」あるいは「きれいな」日本語であり、誤った日本語が処理の入力となることはほとんどない。よって、これらも対象とせざるを得ない日本語校正タスクは非常に困難さが伴い、場合によって形態素解析を行わない、もしくは形態素解析誤りを含むことを前提とした処理が要求される。

さらに、対象とするテキストの性質が非常に多様であり、同様に扱うことができない。まず、執筆者が日本語を母語とするかどうかで誤りの性質は大きく異なり、同じ日本語母語話者であっても子供か一般成人かで分類される。日本語母語話者の大人が執筆した文章であっても、一般的な文章と専門性の高い文章ではチェック項目が異なる。また、日本語学習者であっても漢字圏か非漢字圏かで誤りの傾向が異なることはすでに知られており、これについても同様ではない。

以上のように、「日本語校正タスク」は一見すると同一のタスクのように感じるが、前述したタスクやテキストの多様性を考慮すると同一に議論するのはほぼ困難である。よって、結果的に、**Project Next**

においても万能な日本語校正タスク（の入力テキスト）を定義するのは不可能であり、何らかの形で対象を限定した上でシステムの誤り解析を行うのが現実的と考える。

## 2. 現在の開発状況

次に、前述した多様性の中で、現在の校正／検出システムはどのような執筆者のどのような言語表現を処理対象にしているかについて概略を述べる。

富士通研究所[鄭 2013]： 中国へのオフショア開発向けに誤り事例コーパスを構築。このコーパスに基づいて事例ベース及び校正規則により校正を実現。よって、多種多様な誤りが校正対象となる。

入力文：パラメータの文字列転換はされていない

校正結果：パラメータが文字列変換されていない

東芝ソリューション [熊野 2012]： 機械翻訳への可能性を考慮しつつ、可読性診断機能として構文的な要素の診断機能を持つ。構文的な要素： 曖昧な係り受け、複合語、主語／述語／目的語の省略、主語／目的語と述語が離れている、長い修飾部、述語の数、助詞「は」

検出対象の例：

顧客用文書をカラー、社内用文書を白黒で印刷する。 （名詞で中止）

顧客用文書をカラーで、社内用文書を白黒で印刷する。 （助詞で中止）

顧客用文書をカラーで印刷、社内用文書を白黒で印刷する。 （動詞語幹で中止）

市販ソフトの例（ジャストシステム Just Right! 5 Pro） [Just Right]： 語彙的な校正項目以外としては、助詞関連（助詞抜け、助詞の誤り、同一助詞の連続）、修飾関係、並列関係、呼応表現など

助詞抜け： 学校行く

助詞の誤り： 毎月の懸賞プレゼントとがあります

全ての答えはことわざの中のある。

同一助詞の連続： レストランでゴールドカードで支払う

修飾関係： 白いかごの中の小鳥、大きな黒い目の女の子

並列関係： 太郎および花子または次郎

呼応表現： ちっともおもしろい、決して遅刻する

## 3. 現在の研究状況

日本語校正に関する研究では、主な処理対象は日本語学習者（外国人）が執筆した文章である。この

うち、[水本 2013]のように誤りの種類を限定しない研究もあるが、助詞の誤り検出・訂正のみに限定した研究の報告が最も多い[今枝 2003][南保 2007][大木 2011][今村 2012][笠原 2012]。この他では、前述したように産業日本語向けに富士通研究所が中国へのオフショア開発向けに誤り事例コーパスを構築し、このコーパスに基づいて校正システムの研究開発を行っている[鄭 2012][鄭 2013]。一方日本語教育向けでは、東工大で日本語作文推敲支援システム「ナツメグ」の開発を進めている[八木 2012][八木 2014]。ここでは、レジスターの誤り検出、すなわち日本語として誤っているわけではないが論文やレポートの特定文書では別の表現に書き換えたほうがよい表現の検出を行っている。

日本語話者向けの校正では長岡技術科学大学が保険関連文書向けの校正支援システムの研究開発を進めている[丹治 2011][大平 2012][林 2014]。この研究では、原文書（例：約款）と派生的な文書（例：パンフレット）の比較を行うことで誤りを発見する手法[丹治 2011][大平 2012]や、同音異義語や漢字変換誤りの検出[林 2014]を対象としている。

#### 4. 日本語学習者のテキストに対する誤り訂正の可能性

前節までの議論を踏まえ、本報告では日本語学習者の執筆した文章を対象にして、システムの出力誤り分析ではなく、自然言語処理技術によって誤った日本語がどの程度検出、もしくは訂正可能かについて議論する。

分析は、Web上で閲覧可能なオンライン日本語誤用辞典[誤用辞典]に掲載の例文を対象に行った。同辞典では、誤りを30分類程度（中分類）に分類しており、本分析はこの分類項目ごとにいくつかの例を観察することで分析した。

（助詞）

日本の中で色々な（⇒な）場所に行った。

新しい友達を（⇒が）できたり、自信が増えたりした。

結論だが、ビデオゲームなら（⇒には）様々な良い面もあれば悪い点もある。

概ね対応可能。ただし取り立て詞「は」の用法（特に「が」との使い分け）はn-gramや格フレームでは対応できないので難易度は高い。

（ヴォイス）

その新聞が伝統的に労働党を支持されている（⇒している）から、

結局、母は私に許しました（⇒許してくれました）。

ドーピングのルールをきまる（⇒決める）ことはとても難しくなっている。

一部は常識的（あるいは統計的）な判断で検出可能であろうが、使役や「てくれる」等は非常に困難。

（テンス・アスペクト）

練習する（⇒している）時、私は一度成功しました。

いつもお金を使いすぎました（⇒ます）。

現状では検出困難だが、「練習する時に成功した」とは言わないことから、検出できる可能性はあ

る。同様に、「いつも」は過去形と呼応しないことから検出できる可能性がある。

(基本文型)

また、旅はもちろん楽しいである。

また彼と結婚して、自分の子供つくりたいである。

日本語の選択の原因のは (⇒日本語を選択した理由は) 日本が好きことだ。

当該表現のみで不自然さを感じることから、n-gram 統計等で検出可能。

(基本文型・接辞)

それで、私はラッキーように感じる。

でも、去年のある日、気持はよくないかもしれませんでした。

極東に日本は (⇒日本は極東で) 一番優秀な国です。

概ね n-gram 統計や格フレーム等で検出可能か。(ハを含む) 格要素の出現順も検出できると思われる。

(形式名詞)

夢は何かようになりたい物 (⇒何か実現したいと思っていること) である。

私の夢は日本の家のゆめいな屋根を見ます (⇒見ることです)。

つまり、自分の好きであるφことだ (⇒自分の好きにしているということだ)。

局所的な解決は不可能だが、「夢/原因は~ことです」のような固定性が高いので誤り検出は可能か。

(指示詞)

φ (⇒それは) 今で始まったことではない。

急速に高齢化の問題と言えば、φ (⇒このように) 両親と住むことは大事な原因かもしれない。

φ (⇒その) 一つの例はオリンピックである。

文レベルを超えた談話レベルの結束性を理解する必要があり、おそらく当面は困難。

(疑問詞)

夢は何でしょうφの問題 (⇒何かという問い) に答えて、望みをかなえることだと思う。

外国人はどうしてイギリスがかわいいらしいと思っているφ (⇒のか)。

ある程度語彙的な問題だが、疑問詞(どうして)の場合に文末表現が限定されるのでこの情報を利用すれば検出可能と思われる。

(2語の接続)

現象の支持者も反対者も多くて、いい点と (⇒も) 悪い点も両方あると思う。

両親と (⇒も) 親戚も私は絶対合格だと思っていましたから、私も絶対合格だと思いました。

「AもBも~だ」のような文型を理解すれば概ね対応は可能

(2文の接続)

それから (⇒けれども) 日本語の言語に慣れました。

また (⇒だから)、旅はもちろん楽しいである。

もし、もう一度日本に行けば (⇒行くとしたら)、沖縄に行く。

2文の接続関係は自動でもある程度把握可能である[山本 2008]。誤り検出利用であれば対応可能か。

(修飾)

私も失敗な (⇒失敗した) ことがあります。

そして、イギリスのステレオタイプが多少に事実な (⇒事実と一致する) ところがあるだろう。

また、万一のため、大切 (⇒大切な) 電話番号を調べる。

急速に (⇒急速な) 高齢化の問題と云えば、

多くは語彙的な問題に感じるので検出可能か

(語彙)

でも、私は人前が怖いと思います (⇒です)。

愛する人を傷めるとき (⇒傷つけると)、本当に苦しいです。

これから、私の失敗の経験と言います (⇒を話します)。

バイクは全然 (⇒完全に) 壊れました。

私たちは失敗の中でいい経験を覚えられました (⇒いろいろなことを学びます)。

使用する語彙の選定、及び格助詞の誤用検出は可能

(文全体)

そのため、そのようなゲームに深い感情を感じて (⇒夢中になりすぎて) さらにしてしまうということである。

私の意見では、言語の個人的な資質 (⇒個性) を守るために保守的な立場があるべきである。

世界中、言語を聞くと、別の国の母語が聞える (⇒どの国でも会話の中には外国語が入っている)

私はあくまで自説を主張します (⇒私は必ず夢を実現させたいと思います)。

あとで日本語をべんきょうしたいだとかいけつしました (⇒それで日本語を勉強することにしました)。

昔から言葉はすっかり変わっていることで、その発達は (⇒変わってきたのだから、若者言葉の発達は) 当然だろう

私は多いお金の使わないを知って (多くお金を使ってはいけないと知っていながら)、買います。

「深い感情を感じて」「個人的な資質」の例のように n-gram 情報などでは検出できないものが多数あると予想される。

## 5. 日本語校正システムの出力評価

本節では、異なる会社が開発して運用されている2つの日本語校正システムを用いて、日本語学習者テキストに対して校正処理を行い、その結果に基づき日本語校正技術の現状を議論する。

### 日本語校正システム

本稿で使用した日本語校正システムは、市販製品の校正システム A、および開発企業内のみで使用されている校正システム B である。

校正システム A は、誤字・誤用、不適切な表現や、表記ゆれなどをチェックすることが目的であり、表現の洗練を求めるユーザをターゲットとして開発された。特に新聞社、出版社などの業種のユーザが多く利用しているようである。Table 1 では、システム A の校正項目および日本語学習者テキストに対す

る処理結果の件数を示す。

**Table 1: 校正システムAの校正項目と処理結果件数**

校正項目	説明	検出件数
誤りチェック	誤字脱字、仮名遣い、慣用表現、呼応表現、ら抜き表現、同音語誤り、二重敬語、西暦・和暦など	123
用語基準	送り仮名、漢字基準、公用文、難しい語の言い換え、旧字体、商標・商品名、数字の表記など	207
表現洗練	文体の統一、重ね言葉、同一助詞の連続、二重否定、修飾関係、並列関係など	239
字種統一	単位、句読点、カタカナ、数字、記号、アルファベット	75
長さチェック	文、句読点、ひらがな、カタカナ、漢字	28
環境依存文字	外字と機種依存文字、JIS X 0213:2004 で例示字形が変更された漢字	2
スペルチェック	先頭のみ大文字の単語、すべて大文字の単語、数字を含む単語、全角を含む単語など	5
表記ゆれ	外来語のみ/全部、全角半角・大文字小文字の区別を指摘する/しない	84
括弧	対応、階層	0
印刷標準字体	簡易慣用字体・デザイン差などの文字	0
合計		783

校正システム B は、仕様書などの技術文書の品質を向上させることを目的として開発したものであり、企業内の文書品質を高めるため開発された。以下のような文や文書を作成するための校正を行う。

- 複数の意味に解釈されない文
- 主語や目的語の省略がなく、執筆者以外(翻訳者や外注先)でも内容を正しく理解できる文
- 簡潔・明瞭な文書
- 用語の一貫性のある文書

Table 2 に校正システム B の校正項目および日本語学習者テキストに対する処理結果の件数を示す。

校正システムAと校正システムBは、想定ユーザが異なるため、設計思想も異なっている。Table 1 の校正項目から考察できるように、校正システムAは「きれいな文面」を求める校正を行い、文面の洗練および出版物としての正しさを追究する設計思想である。そのため、文字印刷の際を想定した校正項目が多く見られる。一方、校正システムB (Table 2) では、「表現の正しさ」を求める校正項目を重点的に設計された。企業内の技術文書を想定するため、用語・表現の曖昧さを徹底的に削減することがシステムBの設計思想の特徴である。そのため、文構造の正しさをチェックする項目がシステムAより比較的多い。

本稿では、両システムの利用設定として、全校正項目を出力するように設定した。両システムは共通して、システムのユーザを「日本語話者」と想定している。両システムの校正項目の内容から、両シス

テムは「間違った日本語を正す」ことが目的ではなく、「より良い日本語を追求する」ことが目的であると考えている。すなわち、本稿は両システムを「校正」システムと称しているが、「推敲」システムとしての役目が大きい。また、両システムの出力内容については、指摘箇所は出力するが指摘に対する修正候補を出力できない場合が多い。すなわち、両システムはユーザーに対して「怪しいところ」を指摘し、ユーザーの本当の意図を判断できないため、指摘箇所に対する修正候補を作成できないというのが現状である。そのため、「間違った日本語」に対する校正は本来の設計ではない利用シーンのため、校正効果は本来の設計性能を発揮できないと考える。

**Table 2: 校正システムBの校正項目と検出件数**

校正項目	説明	検出件数
言葉使いのチェック	同義語、非専門用語、未登録語、受け身の使用を指摘	202
動作処理のチェック	動詞+「行う」、感情表現（～したい）、使役表現の使用を指摘	30
あいまい表現のチェック	否定表現の並立、二重否定、部分否定、接続助詞「ため」、格助詞「より」の使用を指摘	48
内容複雑さのチェック	接続助詞「が」、「ので」の使用を指摘、連用中止の箇所を分割するように推奨	93
修飾関係のチェック	修飾先が複数あり、あいまい性がある修飾関係を指摘	3
処理対象のチェック	目的語の省略、形容（動）詞・副詞の使用、指示代名詞の使用による処理対象があいまいになる事例を指摘	459
助詞チェック	「てにをは」などの助詞誤用を指摘	43
文体チェック	「である」、「ですます」体の不統一を指摘	149
合計		1027

### 日本語学習者テキストと処理結果

本稿での出力評価の処理対象は、公開されている「オンライン日本語誤用辞典」の事例を元に収集した日本語学習者テキストである[誤用辞典]。公開サイトから検索可能な全事例 1023 件を収集した。同一例文には複数の校正箇所があるため、全事例に対してさらに整理し、校正前の原文 491 文を抽出した。抽出した 491 文を処理対象として評価システムに入力した。本処理対象のテキストは日本語学習者の事例であるため、収集したテキストは「間違った日本語」が多いと考えられる。前節で議論したように、学習者テキストは評価用の両システムとして処理が困難の可能性はある。

Table 3は日本語学習者テキストの誤用分類と件数を示す。元データの誤用分類は三段階の詳細定義があるが、Table 3は大分類（文法、語彙、句・文全体）の件数のみを記載する。表に示すように、大分類では約6割（652件）が文法の誤用であり、語彙の誤用は約3割（334件）である。おそらく、学習者の作文において学習者はすでに勉強した言葉（自信がある語彙）の意味を正確に選択・利用できるが、

文法を利用することには経験が少ないため、文法の誤用が多くなる。一方、Table 1とTable 2の校正項目から、校正システムAとBはいずれも語彙についての表記の統一と洗練に注力した設計であることが分かる。

**Table 3: 日本語学習者テキストの誤用分類（大分類）と件数**

誤用分類(大分類)	説明	件数	システムAの一致件数	システムBの一致件数
文法	助詞、複合辞、文型、テンス・アスペクトなどの誤用例	652	10	3
語彙	動詞、形容詞、名詞、副詞、連体詞、接辞、連語などの誤用例	334	97	65
句・文全体	文（句）の意味が不適切のため、全体的に書き直す修正	37	0	0
合計		1023	107	68

### 処理結果に対する考察

Table 3の「システムAの一致件数」と「システムBの一致件数」には、校正システムAと校正システムBの出力結果が原文の誤用指摘箇所と一致する件数を示す。我々は校正システムAと校正システムBの出力結果自体を評価しない。両システムの設計思想と学習者テキストの性質が異なるため、両システムの出力は必ずしも日本語教育における校正にあてはまるものではない。このことから、我々は両システムの検出結果と学習者テキストの元の指摘とが一致する件数のみを検証した。

Table 3の結果から、全件数に対する両システムの指摘数は1割未満であることがわかった。誤用分類として両システムの結果はほとんど語彙の誤用分類である。ほとんどのシステム指摘は言葉表記の誤用である。両システムでは、タイプミスや言葉の誤りに対し、辞書の未登録語として解析する結果になる。また、両システムは、助詞誤用の検出機能を持つため、助詞の誤用に対する指摘も少し成功件数がある。

校正システムAの出力結果では、表現洗練、用語基準と誤りチェックの校正項目の件数が多い。特に文体（です/である体）、送り仮名、漢字基準と公用文表現の指摘が多かった。一方、校正システムBの出力には、処理対象のチェックと言葉使いのチェックの校正項目の件数が多かった。特に技術文書では望ましくない感情・修飾表現が多く指摘された。両システムはいずれも設計思想に基づいた校正結果を出力したが、日本語の「誤り」ではない指摘が多いことが分かる。

一方、Table 3の結果から、両システムは共に文法の指摘がほとんど出来ていない。設計上、両システムはある程度の文法誤用（助詞などの誤用）を指摘することができるが、日本語学習者テキストに対してその性能が発揮出来ていない。日本語学習者テキストの原文数（491文）とその指摘合計数（1023件）から察知できるように、多くの入力について複数の誤用指摘がされている。校正システムの想定ユーザは高い日本語能力を持つ話者（または日本語母語話者）ため、形態素解析などの言語処理技術が解析可能の前提で校正を行う設計であると推測できる。文法と語彙の誤用が同時に出現する場合、辞書未



登録語の出現とともに形態素解析が失敗になるため、文法の校正が困難になる。そのため、文法に関する校正システムの指摘性能が低下する。

## 6. まとめ～日本語校正のために何を研究すべきか～

これまで述べてきたように、究極的な日本語校正実現のためには日本語解析のほぼ全要素が必要とされ、日本語校正技術の更なる進展には日本語処理全体の底上げが必要である。ここでは、これらのうち従来あまり検討されておらず、かつ我々が重要であると考える2項目を指摘する。

(A) 誤った日本語に対する形態素解析技術

(B) 分離表現（例：～しか～ない、いつも～する）の把握

まず、今回使用したシステムがどちらも形態素レベルでの入力誤りを想定していないことから分かるように、この対処は必須かつ重要である。誤った日本語の形態素解析は日本語校正だけでなく日本語教育やネットスラングの解析などにおいて需要が大きい、これに対する研究は[藤野 2012]など極めて少ない。このため、今後の進展が必要であるが、特に機能が限定的であってもツール化や公開が関連技術全体の進展に必要と考える。

次に、4節および5節で指摘したように、現状のシステムは語彙的な誤りの指摘が中心であり、文法項目の誤検出についてはまだ十分とは言えない。特に分離して表現される「～しか～ない」「～したり～したりする」「いつも～する」などの表現については、現状で言語資源が何一つ整備されているとは言えない状況である。局所的な表現を文節や複合的機能辞などにまとめあげる研究（チャンキング）は従来行われてきたが、このような研究だけでは分離表現の把握が困難である。よって、まずはこのような表現の言語資源化を進め、次に何らかの表現方法でまとめあげる処理が必要だと考える。

## 参考文献

- [今枝 2003] 今枝 恒治, 河合 敦夫, 石川 裕司, 永田 亮, 榊井 文人. 日本語学習者の作文における格助詞の誤り検出と訂正, 情報処理学会研究報告, コンピュータと教育研究会, Vol.2003, No.13, pp.39-46, 2003.
- [今村 2012] 今村 賢治, 齋藤 邦子, 貞光 九月, 西川 仁. 小規模誤りデータからの日本語学習者作文の助詞誤り訂正. 自然言語処理, Vol.19, No.5, pp.381-400, 2012.
- [大木 2011] 大木 環美, 大山 浩美, 北内 啓, 末永 高志, 松本 裕治. 非日本語母国話者の作成するシステム開発文書を対象とした助詞の誤用判定. 言語処理学会第 17 回年次大会発表論文集, pp.1047-1050, 2011.
- [大平 2012] 大平 真一, 山本 和英. 保険関連文書を対象とした校正支援システム. 言語処理学会第 18 回年次大会発表論文集, pp.243-246, 2012.

- [笠原 2012] 笠原 誠司, 藤野拓也, 小町 守, 永田 昌明, 松本 裕治. 日本語学習者の誤り傾向を反映した格助詞訂正. 言語処理学会第 18 回年次大会発表論文集, pp.14-17, 2012.
- [熊野 2012] 熊野 明, 加納 敏行. 機械翻訳精度を向上させる可読性診断技術. JAPIO Yearbook, 2012, pp.280-285, 2012
- [誤用辞典] オンライン日本語誤用辞典 (公開版 Ver.1.1) , 東京外国語大学望月圭子研究室  
[http://cblle.tufs.ac.jp/lc/ja\\_wrong/](http://cblle.tufs.ac.jp/lc/ja_wrong/)
- [Just Right] <http://www.justsystems.com/jp/products/justright/>
- [丹治 2011] 丹治 広樹, 山本 和英. 保険約款と派生書類の自動対応付け. 言語処理学会第 17 回年次大会発表論文集, pp.868-871, 2011.
- [林 2014] 林 秀治, 山本 和英. 保険関連文書を対象とした文章校正支援のための変換誤り検出. 言語処理学会第 20 回年次大会発表論文集, pp.618-621, 2014.
- [藤野 2012] 藤野 拓也, 水本 智也, 小町 守, 永田 昌明, 松本 裕治. 日本語学習者の作文の誤り訂正に向けた単語分割, 言語処理学会第 18 回年次大会 発表論文集, pp. 26-29, 2012
- [水本 2013] 水本 智也, 小町 守, 永田 昌明, 松本 裕治. 日本語学習者の作文自動誤り訂正のための語学学習 SNS の添削ログからの知識獲得, 人工知能学会論文誌, Vol.28, No.4, pp.420-432, 2013.
- [南保 2007] 南保 亮太, 乙武 北斗, 荒木 健治. 文節内の特徴を用いた日本語助詞誤りの自動検出・校正. 情報処理学会研究報告, 自然言語処理研究会, Vol.2007, No.94, pp.107-112, 2007
- [八木 2012] 八木 豊, ホドシチェク・ボル, 仁科 喜久子. BCCWJ と学習者作文コーパスを利用した日本語作文支援—表記と共起に関する誤用添削プロトタイプ構築—. 第 1 回コーパス日本語学ワークショップ予稿集, pp.315-320, 2012
- [八木 2014] 八木 豊, ホドシチェク・ボル, 阿辺川 武, 仁科 喜久子. 日本語作文推敲支援システム「ナツメグ」における誤用検出手法の評価. 第 5 回コーパス日本語学ワークショップ, pp.167-170, 2014
- [山本 2008] 山本 和英, 齋藤 真実. 用例利用型による文間接続関係の同定. 自然言語処理, Vol.15, No.3, pp.21-51 , 言語処理学会, 2008.
- [鄭 2012] 鄭 育昌 長瀬 友樹. 外国語母語話者が作成する日本語技術文書を対象とした訂正履歴の分析. 言語処理学会 第 18 回年次大会 発表論文集, pp.34-37, 2012
- [鄭 2013] 鄭 育昌 長瀬 友樹. オフショア開発向けの事例ベース日本語自動校正システムの構築. 言語処理学会 第 19 回年次大会 発表論文集, pp.476-479, 2013