

ICTを活用した非母語話者向け日本語学習支援システムの とりのくみ

稲積宏誠^{†1} 大野博之^{†2}

著者らは、情報工学と自然言語処理技術に基づく母語話者のための日本語校正・推敲ツールの開発を進め、現在非母語話者向けへと展開している。これは、公開済み日本の辞書を利用し、任意の文章が基本文型に合致した構造（助詞の用い方）となっているか否かを判定するものである。本稿では、その概要と問題点を述べる。

Developing ICT-based Learning System to Enhance Non-Native Learners' Understanding of Japanese

HIROSHIGE INAZUMI^{†1} HIROYUKI OONO^{†2}

The authors have developed the tool for proofreading and elaborating Japanese sentences, which is as for native speakers. They have been continuing this project, and then started to expand the tool for non-native speakers. Summary of the tools currently under development are as follows. (1) Use Japanese dictionary published free for educational and research use. (2) Determine whether or not any text written by students is the correct syntax by matching the basic sentence patterns with the dictionary. In this manuscript, we describe overview and problems.

1. はじめに

著者らは、2007年から、情報工学と自然言語処理技術に基づく母語話者のための日本語校正・推敲ツールの開発を行ってきた。その背景は、近年の母語としての日本語表現力の低下であり、特に正確で簡潔な表現による論理的な文章を作成する能力の低下である。ここで言う日本語表現力とは、事実を客観的に示し、明確に主張を述べることができ、また、複数の解釈を与えず、冗長な表現を含まない文章を作成することのできる表現力とする。すなわち、レポート、論文、マニュアル等の技術文書などの作成に求められる素養に相当する。

日本語校正・推敲ツールの開発においては、論理的な展開となるような文章表現上のシンタックスを満足した表現力を狭義の論理的な文章と位置づけ、その育成を第一義とした。その結果、文章表現上のシンタックスを効率的に学習できる環境の構築により、論理展開のセマンティクス学習への人的資源の配置を十分可能となる。このことが、一貫した本研究の目指す教育分野への貢献である。現在開発している校正・推敲ツールは、フリーの日本語解析ソフトを用い、独自開発の日本語誤り検出アルゴリズムと関連辞書を組み込むことにより、Microsoft Windows 環境下のフリーソフトとして実現・公開されている。[1]

現在、母語話者向けの開発により得られた知見に基づき、非母語話者のための学習支援ツールの開発を開始している。

非母語話者向けのツール開発にあたっては、基本文法事項についても、その理解を促す環境が必要となる。ところが、母語話者向けの支援ツールは、日本語の基本文法事項を理解していることを前提としているため、このような基本文法事項は考慮に入れていなかった。したがって、この点を非母語話者への取組の最優先項目とし、まず動詞・形容詞に対応する助詞の用い方に注目した基本文型の理解を目的とした学習システム構築を行うこととした。

基本文型の学習には、教育・研究用に公開されている（独）情報処理推進機構(IPA)（現在は特定非営利活動法人 言語資源協会(GSK)が管理）が開発した日本語辞書 IPAL (IPALexicon of the Japanese Language for computers) を用いる。開発されてから年月が経過し、その後いくつかの辞書が整備されたものの、広く利用可能な教育用ツールを開発するためには、その使用にあたっての制約が大きい。その結果、IPALが研究・教育用に利用できる数少ない辞書のなかで最も利用価値が高いと判断した。IPALには、語彙体系上・使用頻度上重要であると考えられる基本的な動詞(861語)、形容詞(136語)、名詞(1081語)が収められている。それらは、意味・統語的な特徴に基づいて区分し、形態、意味、統語、慣用表現などに関する情報が詳細に記載されている。

現在開発中のプロトタイプシステムでは、任意の文章に対して IPAL で分析可能な動詞、サ変動詞、形容詞、形容動詞が含まれているか否かを判定し、その文章の係り受け構造を表示し、基本文型に合致した構造（助詞の用い方）となっているか否かを判定することを可能にしている。なお、日本語学習支援システムとしては2つの利用形態を想

^{†1} 青山学院大学
Aoyama Gakuin University.
^{†2} 東京医療保健大学
Tokyo Healthcare University

定している。一点目は、正誤問題、語の補充問題などの問題の解説を行うことを目的とした自学習および授業支援である。二点目は、日本語文章作成時の校正・推敲支援を目的としたものである。

本稿では、まず母語話者向けに開発している構成・推敲ツール Tomarigi の概要を説明し、このツールを用いた非母語話者向け日本語学習支援の可能性を論じる。次に、IPAL 辞書を利用した非母語話者向けツールの概要を説明する。最後に、これらの問題点と今後の展開を論じる。

2. 母語話者向け校正・推敲ツールの概要

日本語を対象とした自然言語処理研究は、現在、形態素解析ツールである Mecab[2]・JUMAN[3]や、係り受け解析ツールである Cabocha[4]・KNP[5]に代表される基礎研究の成果によって、様々な応用研究が可能となった。例えば、日本語エディタ、機械翻訳、テキストマイニングなどが代表的な応用例である。ただし、その処理量の大きさから、文章作成の補助としての校正支援や推敲支援を含むインテリジェントエディタの開発には、限界があるとされてきた。そのような中で、推敲支援を実現する試みも進められ、教育関係への展開にも関心が及びつつある。

教育関係への展開は、主として第二言語として学ぶ学習者の読解および作文支援としてとして取り上げられることが多い。代表的なシステムである読解支援システム「あすなる」[6]は、Web 上で学習可能な第二言語読解学習支援を多言語対応システムとして開発されたもので、学習者の母語による支援により文章を理解できることを目指している。

一方、作文支援システムとしては、「あすなる」の開発者による「なつめ」[7]をはじめとする第二言語学習者向けの取組に加えて、より高度な文章推敲や校閲を支援するための取組も見られるようになった。これらの研究では、作文教育において、作文とそれに付随する添削結果をデータベースに蓄積し、教師の誤用分析などを支援している。

これらを踏まえ、著者らは、テキストエディタとして実際に文章を作成しながら、校正・推敲に関する機能を利用できる校正・推敲ツールの開発を開始した。[8] また、これらの校正・推敲機能は、機能自体を別ファイルで作成することで任意に追加することが可能な点を特長とする。また、最大の利点は、校正・推敲機能を活用することにより演習問題の自動生成機能を実現することが可能となることである。この取組についても継続させているが、なお実用に向けての課題が多く、公開には至っていない。この詳細については別の機会に発表することとする。

校正・推敲ツールは、図1のように、文書の編集を行うエディタ領域、校正・推敲対象の指摘カードおよび指摘詳細カード、さらに単語数、異なり語数などの文章全体の客観的評価指標を示す文書プロパティ領域から構成されてい

る。これは、「書くこと」と「指摘を受けること」と「修正すること」を明確に分けて、ユーザーに意識づけるためである。

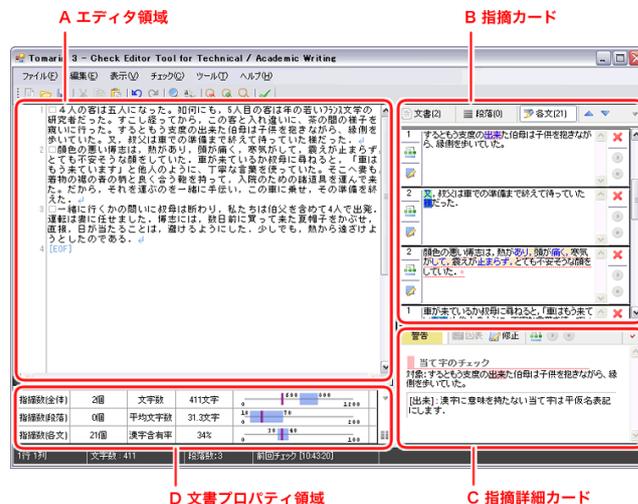


図1 母語話者向け構成・推敲ツール

Figure 1 The tool for proofreading and elaborating Japanese sentences for native speakers.

基本的な校正機能としては、レポート・論文を想定した「体言止め」「文末の無駄な表現」「平仮名/漢字書きすべき副詞」「全角英数字と半角カタカナ」「漢数字」「読点」「常用漢字外使用」「接続詞の漢字使用」「平仮名書きすべき形式名詞」「不適切な表現」「句点の統一」「文末(である・ですます調)」「当て字」などのチェック機能が実装されている。

一方、推敲支援では、表記上の誤りではなく、理解容易性の向上や読みやすさの改善を目的に文を訂正することを目的としている。現在は3つの推敲支援機能を実装している。

前置修飾節と長文の判定: 二重の前置修飾節を持つ文と長文を典型的な悪文と位置づけ、その検出を行い、指摘を行う。二重の前置修飾節を持つ文とは、修飾節の中に修飾節があるような文である。一方、長文とは、ここでは人の言語理解過程に基づいて定めたものとする。すなわち、その言語理解過程において、形態素解析や構文解析などの処理が行われているという前提で、これらの過程で作られる処理結果が短期記憶に保持できる容量を超えるか否かの判定を行うものである。これらは、係り受け解析の結果から、「節」の判定を行うことで、二重の前置修飾節を抽出することができる。また、1文節を記憶単位としたうえで、連体詞や節として解釈した場合は、記憶単位に統合するなどの処理をしながらカウントし、設定した短期記憶容量と比較して、長文の判定を行うことを可能にしている。

読点の位置: 文中で使用される読点は、主に、文意に誤解を与えないようにすることと、文を読みやすくすることを目的として使用される。この際、動詞の連用形の後には

読点を打つべきであり、隣接する文節同士に係り受け関係がある場合には打つべきではないことが知られている。ただし、名詞が羅列されている場合は、隣接している場合でも、その区切りに読点を打つ必要がある。そこで、係り受け解析の結果を利用して、これらの基準をもとに、付与すべき読点および除去すべき読点についての指摘を行う。

係り受け解釈のあいまい性：係り受け解析ツールによって、最も尤もらしい係り受け関係を求めることができる。そこで、係り受け関係に基づいた文節単位での構造を視覚化することによって、何が解釈の違いを導くのかについて検討することを可能とする。これによって、構文上の特徴や修飾語と非修飾語の距離による解釈の違いや文の曖昧性を容易に確認することができる。

3. 非母語話者向け日本語学習支援システムのとりくみ

本稿で取り上げる学習支援は、学習者自身が文章作成時に自らその誤りへの気づきを促すこと、また日本語教師らの教育実践における支援を行うことを目的とする。これは、著者らの母語話者に対するこれまでの取り組みのなかで、単に文章上の誤りを発見し、訂正をさせるのみでは学習効果がないとの知見を得ているからである。

3.1 母語話者向けツールの活用

著者らは、まず母語話者向けに開発した校正・推敲ツールを非母語話者向けに展開することを検討した。[9][10]

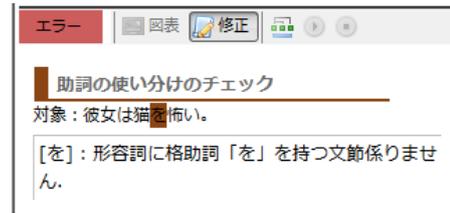
母語話者向けの学習支援ツールを非母語話者向けのものに展開するためには、非母語話者の犯す誤用の分析と誤用発見ルールの構築が必要となる。誤用については、多くの分析例が示されているが、それぞれの母語との関連性や背景となる文化の違いもあり、文章作成者の意図に沿った修正の実現は困難な課題とされている。[11]

誤用分析そのものについての詳細な検討を行う前に、まず学習システムとしての可能性を探ることとした。誤用の典型例として「助詞」に注目し、そのいくつかを例にとって、誤用パターンとそのチェックルールを試験的に構築し実装した。実装した誤用は、①形容詞の係り受け元に助詞「を」を用いる。②名詞・形容動詞語幹に助詞「の」を用いる。③動詞あるいは動詞型以外の助動詞に助動詞「べき」を接続させる、の3種類である。

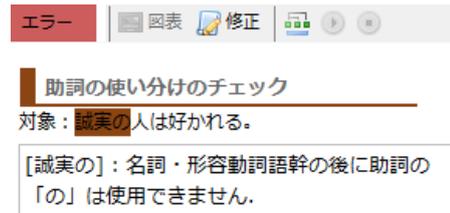
次に、「係り受け関係」を表示し、文の構造を示すことで、文の示す意味と書き手の意図の違いを気づかせることを試みる。非母語話者の陥りやすい誤りの一つに、助詞「は」と「が」の主語としての使い方が挙げられる。この場合、従属節の主語としてそれぞれを用いることで、「係り受け関係」が変わり、文意が変化することがわかる。

以上の4種類の取組例を図2に示す。特に、係り受け関係については、従属節の推定機能により、文意の変化が示

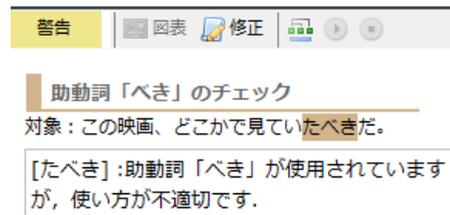
され、動詞「受ける」の主語として用いる場合には「は」ではなく「が」としなければならないことが確認できる。



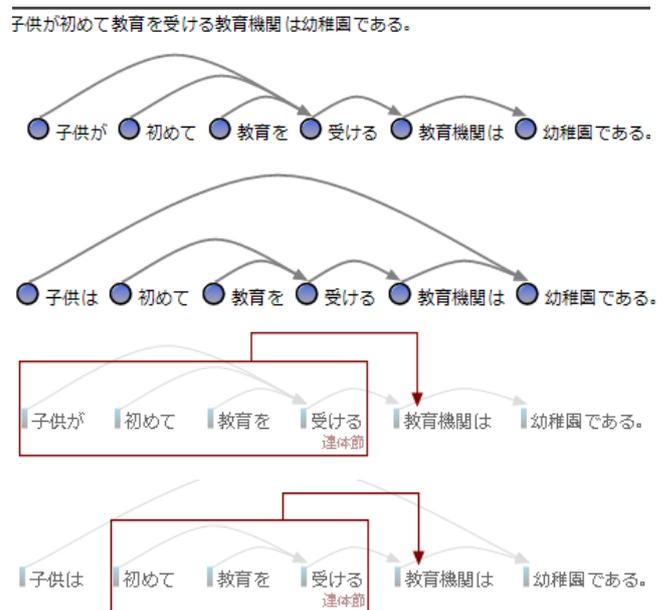
(a) 「彼女は猫を(が)怖い」へのチェックルール適用例



(b) 「誠実の人は好かれる」へのチェックルール適用例



(c) 「この映画、どこかで見ていたべきだ」へのチェックルール適用例



(d) 「子供は/が初めて教育を受ける教育機関は幼稚園である」の係り受けによる気づきの例

図2 母語話者向けツールによる非母語話者の誤用チェックの実装例

Figure 2 Examples of Checking misuse of sentence written by non-native speakers using the tool for proofreading and elaborating Japanese for native speakers.

以上の例で示されるように、形態素解析および係り受け解析結果に個々の誤用判定ルールを適用させることが可能であることがわかる。しかし、誤用パターンを詳細に分析し、一つ一つの判定ルールを構築するのは非常に困難な作業となる。なぜならば、書き手の意図の推定を伴うからである。また、シソーラスなどを含む高度な辞書機能を必要とすることも予想される。例えば、自動詞と他動詞の区別なども、独自辞書を構築しなければ、その判定は実現できない。

そこで、この取組については継続させながら、従来の取り組みを活かしつつ、非母語話者に特化した学習支援の検討を開始することとした。

3.2 基本文型理解のための日本語学習システム

非母語話者に特化した学習支援環境の構築については、日本語教育分野における教育論や誤用分析についても考慮する必要がある。しかし、教育方法論の変遷はあるものの、語彙や文法、基本文型の理解については、その必要性は定着している。そこで、前述した日本語辞書 IPAL に着目し、基本動詞・形容詞・形容動詞といった「述語」を中心とした基本文型の理解とそれに伴う助詞の使い方に特化したシステム構築を行うこととした。

日本語辞書 IPAL (IPALexicon of the Japanese Language for computers) では、語彙体系上・使用頻度上重要であると考えられる基本的な動詞、形容詞、名詞が収められており、意味・統語的な特徴に基づいて区分し、形態、意味、統語、慣用表現などに関する情報が詳細に記載されている。これらの詳細情報をどのように活用するかという点が第一の検討課題となる。そこで、まず基本文型の理解に限定したプロトタイプシステムを検討することとした。

プロトタイプシステム構築にあたり、まず、動詞、サ変動詞、形容詞、形容動詞に関する詳細情報から、「読み」と「基本文型」および「基本文型に対応する意味」を抜き出し、データベースを構築した。また、基本文型では、「助詞」の取り扱いに加えてとりうる名詞の素性にかかわる情報が付加されている。IPAL の提供する名詞辞書と連携することで、より正確な構文の推定が可能となるはずである。データベースにおける記述の抜粋は以下のとおりである。

WV03329 わかれる 別れる ガ:HUM,ト:HUM
 WV03330 わかれる 別れる わかれさす ガ:HUM,カラ/ト:HUM
 WV03331 わかれる 分れる,分かれる わかれさす,わけ ガ:HUM
 WV03332 わかれる 分れる,分かれる ガ:ABS,ニ:ABS
 WV03333 わかれる 分れる,分かれる ガ:CON,カラ/ト:CON
 WV03334 わかれる 分れる,分かれる ガ:CON,ニ:ABS
 WV03335 わかれる 分れる,分かれる わける ガ:CON,ニ:CON/ABS/QUA

 WV03329 一緒にいた者と別々になる。
 WV03330 家族関係を絶って別々になる。
 WV03331 いくつかの小さな集団になる。
 WV03332 複数の人の考えがいくつかに分散する。
 WV03333 本体から離れて、別の物として成り立つ。
 WV03334 一つの物が二つ以上に分岐する。
 WV03335 一つの物が、複数の小部分に区切ったり分類したりすることができる

 AA00983 (ほしい,欲しい) ガ:HUM,ガ/ヲ:***
 AA00984 (ほしい,欲しい) ガ:HUM,ニ:LOC,ガ:CON
 AA00985 (ほしい,欲しい) ガ:HUM,ニ:HUM/ORG/STA,ガ:***
 AA00986 (ほしい,欲しい) ガ:HUM,ニ:KND,ガ:CON
 AA00987 (ほしい,欲しい) ガ:HUM,ガ/ヲ:HUM,ニ:KND

AA00983 (私は)何かを手に入れたと思う。
 AA00984 (私は)どこかに何かがあることを望んでいる。
 AA00985 (私は)どこかに何かがあることを望んでいる。
 AA00986 (私は)どこかに何かがあることを望んでいる。
 AA00987 (私は)どこかに何かがあることを望んでいる。

AV00117 りっぱ 立派 ガ:HUM
 AV00118 りっぱ 立派,りっぱ ガ:CON/APP/INF/PRC,*:
 AV00119 りっぱ 立派,りっぱ ガ:CON,ガ:CON
 AV00120 りっぱ 立派,りっぱ ガ:HUM,ガ:MAN

AV00117 行いや人格などが賞賛に値する。
 AV00118 賞賛に値するほど素晴らしい。
 AV00119 賞賛に値するほど素晴らしい。
 AV00120 賞賛に値するほど素晴らしい。

また、各名詞の意味素性は表 1 のようにまとめられており、「基本文型」のなかに組み込まれている。

表 1 名詞の意味素性

Table 1 Semantic Features of Subjective Nouns

CON	concrete	具体名詞	
ANI	animal	動物	犬馬鳥猿コアラ
HUM	human	人間	姉先生男性学生
ORG	organization	組織・機関	国企業警察研究所
PLA	plant	植物	花桜松バラ
PAR	parts	生物の部分	頭足腕腹根羽
NAT	natural	自然物	山空石川丘
PRO	products	生産物・道具	紙車パン布鉄
PHE	phenomenon	現象名詞(自然/...	光音火風雨涙声匂い病気しわ
ABS	abstract	抽象名詞	
ACT	action	動作・作用	勉強練習見学散歩
MEN	mental	精神	心意識思い出悩み
LIN	linguistic products	言語作品	名前ニュース説教
CHA	character	性質	美欠点みかけ寛容
REL	relation	関係	縁原因条件根拠
LOC	location	空間・方角	外公園東右
TIM	time	時間	昨日.日曜日.夕方
QUA	quantity	数量	3日.3人.5kg.4m.全部.一人ずつ
DIV	diverse	制限緩やか	

これらの情報に基づいて構築したプロトタイプシステムのインタフェースを図 3 に示す。

左上から、対象文編集ウィンドウ、係り受け関係表示ウィンドウ、右上から述語(動詞、サ変動詞、形容詞、形容動詞)基本文型評価ウィンドウから構成されており、必要に応じて形態素情報および詳細な係り受け関係の評価を行うウィンドウ、名詞の素性情報を示すウィンドウを起動させることができる。

図 3 では、「その時日本の友人の家で住みました」が評価対象の文となっている。チェックを行うことで、動詞「住む」に関する評価を行うことができる。その結果、もし「ある場所に居を構えて住む」という意味なのであれば、基本文型は「○○が××に住む」であることが示され、「××で」という表現は不適切であると評価される。また、それぞれの名詞素性が意図しているものかどうかも同時に確認することができる。例えば、「家」が空間や方角を意味する名詞と判断されれば、この基本文型に合致するとみなすことができることになる。また、形態素解析の結果についても参照可能である。



図3 非母語話者向け日本語学習システムのプロトタイプ
 Figure 3 The prototype : ICT-based learning system to enhance non-native learners' understanding of Japanese

次に、「父は日本の興味がありました」と「父は日本に興味がありました」の2文について評価した例を図4, 5に示す。

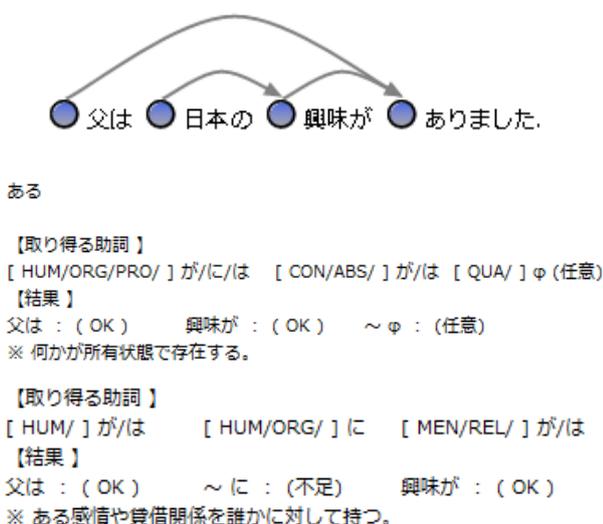


図4 プロトタイプシステムの利用例1
 Figure 4 Example 1 of the use of the prototype system

ここでは、「ある」という動詞の基本文型を確認することになる。前者については、「何かがある状態が存在する」には合致するが、「ある感情や貸借関係を誰かに対して持

つ」の意味には合致しないことが、また後者については、「何かに関してある繋がりなどが存在する」という意味として解釈も可能であることが示される。さらに、名詞の意味素性からは、「日本」が抽象名詞ではないことから、形式的に同じ文型であっても、その意味する内容の違いから絞り込むことが可能となる。

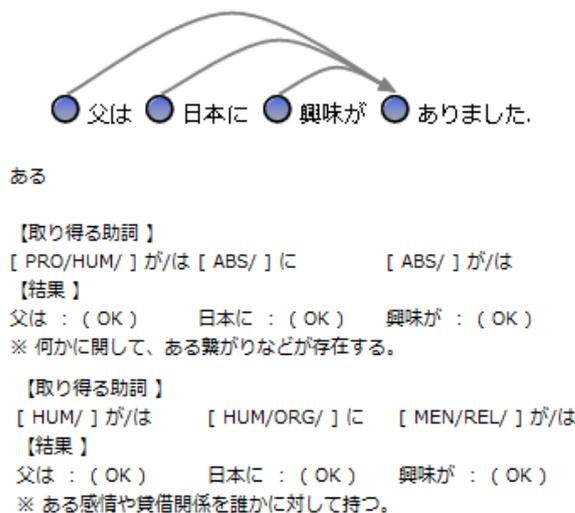


図5 プロトタイプシステムの利用例2
 Figure 5 Example 2 of the use of the prototype system

4. おわりに

本稿では、母語話者向けに開発を続けてきた校正・推敲ツールを非母語話者向けに拡張する取組と、非母語話者向けに特化した日本語学習支援システムの考え方を示し、プロトタイプシステムを紹介した。

日本語文章の解析においては、形態素解析や係り受け解析の精度が向上したことから、辞書や誤用判定ルールの構築により、非母語話者に対する学習支援の可能性は高まっていると考えられる。今回は、多く用いられている動詞、サ変動詞、形容詞、形容動詞の基本文型に基づき、「助詞」の誤用についての学習支援の可能性を示すことができた。

今後は、システムの精度向上、ユーザインタフェースの改善などを図り、実用に供することのできるシステムへより近づけるための取り組みを継続させていく予定である。

謝辞 本取組は、一部、私立大学共済事業団学術振興資金の援助を受けて進められている。ここに謝意を表す。

参考文献

1) <http://www.pawel.jp/>
 2) Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto(2004) : Applying Conditional Random Fields to Japanese Morphological Analysis.

Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237

- 3) 黒橋禎夫, 河原大輔(2009): 日本語形態素解析システム JUMAN version 6.0 使用説明書. 東京大学大学院情報理工学系研究科
- 4) 工藤拓, 松本裕治(2002): チャンキングの段階適用による係り受け解析. 情報処理, Vol 43, No. 6, pp.1834-1842
- 5) 黒橋禎夫, 河原大輔(2009): 日本語構文解析システム KNP version 3.0 使用説明書. 東京大学大学院情報理工学系研究科
- 6) 阿辺川武, 八木豊, 戸次徳久, 澤谷孝志, 奥村学, 仁科喜久子, 杉本茂樹, 傅亮(2003): 「日本語学習システム「あすなろ」開発の新しい展開 - 構文学習とその評価 -」情報処理学会第 65 回大会論文集 3T2-6
- 7) 阿辺川 武, Hodoscek Bor, 仁科 喜久子(2010): 日本語作文支援システム「なつめ」における共起語検索方法の改善..特定領域研究「日本語コーパス」平成 21 年度公開ワークショップ予稿集. 243-244
- 8) 大野博之・稲積宏誠(2008): “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”. 電子情報通信学会技術研究報告, 教育工学研究会, pp.31-36
- 9) 稲積宏誠, 大野博之(2012): 基本文型理解のための ICT を活用した日本語学習システムの開発: 基本設計第 9 回マレーシア日本語教育研究発表会資料
- 10) 稲積宏誠(2012): 基本文型理解のための ICT を活用した日本語学習システムの開発: プロトタイプシステム第 9 回国際日本語教育・日本研究シンポジウム論文集
- 11) 市川保子(2010)編: 日本語誤用辞典 外国人学習者の誤用から学ぶ日本語の意味用法と指導のポイント, スリーエーネットワーク