社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

EscortJacs: 外国人による日本語文作成を支援する用例文検索システム

加藤　宏紀†　葛原　和也††　加藤　芳秀†††　松原　茂樹††

† 名古屋大学工学部，〒 464-8601 名古屋市千種区不老町
†† 名古屋大学大学院情報科学研究科 〒 464-8601 名古屋市千種区不老町
††† 名古屋大学情報基盤センター 〒 464-8601 名古屋市千種区不老町

あらまし　本論文では，日本語文の作成支援を目的とした用例検索システム EscortJacs を提案する．外国人による日本語文の作成では，特に助詞の誤用が多いことが知られており，本研究では助詞に着目した用例検索を対象とする．EscortJacs では，係り受け関係を考慮することにより，高い精度での検索を実行する．また，係り受け構造，及び，付属語を考慮した分類を行うことにより，ユーザが適切な用例文を容易に発見可能な提示を実現する．評価実験を行った結果，ベースラインと比べ，検索性能，及び，分類性能が向上することを確認した．

キーワード　情報検索, 係り受け構造, 分類, コーパス, 助詞

EscortJacs: Example Sentence Search for Supporting Japanese Composition by Foreigners

Hironori KATO†, Kazuya KUZUHARA††, Yoshihide KATO†††, and Shigeki MATSUBARA††

† Faculty of Engineering, Nagoya Univ. Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan
†† Graduate School of Information Science, Nagoya Univ. Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan
††† Information Technology Center, Nagoya Univ. Furo-cho, Chikusa-ku, Nagoya 464-8601 Japan

Abstract　In this paper, we propose an example sentence search engine for supporting Japanese composition, EscortJacs. In this study, we focus on example sentence retrieval based on particles, since it seems foeigners mistake them in Japanese composition. EscortJacs executes sentence retrieval with high precision by consideing dependency relations. To show the effectiveness of our system, we conducted the experiment. In this result, we confirmed the effectiveness of our system from the viewpoints of search performance.

Key words　information retrieval, dependency structure, classification, corpus, particle

1. ま え が き

近年，多くの留学生が日本に滞在し，日本語教育を受けている．しかし，留学生にとって正しい日本語文を書くことは難しく，留学生を対象とした日本語文の作成支援環境が求められる．そのような環境として，用例検索システムが有用である [1, 2]．用例検索を用いることにより，ユーザはキーワードを入力すれば，文を作成する上で参考になる用例文を参照できる．

本論文では，日本語文作成支援を目的として用例文検索システム EscortJacs を提案する．留学生の日本語文作成において，助詞の誤りが多いことが知られており [3]，助詞の使い方を学ぶことができる用例検索の実現が望まれる．

EscortJacs は，日本語キーワード系列を入力として受け取り，キーワード同士を接続する適切な助詞を明らかにするような用例文を提示する．キーワード同士を接続する関係としては係り受け関係を利用し，キーワードを含む文節間に直接的な係り受け関係が存在する文のみを用例文として提示する．これにより，不要な用例文を検索結果から排除できる．また，用例文を提示する際には，出現するキーワードが構成する係り受け構造，及び，キーワードに後続する付属語に基づき用例文を分類する．これにより，ユーザは，助詞の使用法を区別して用例文を調べることができる．

係り受け関係を用いた検索と分類の有効性を確認するために評価実験を行った．検索性能の評価実験において，提案手法の検索精度は単純なキーワード検索と比べ 20.4 ポイント向上し，係り受け関係を用いることの効果を確認した．また，分類性能の評価実験においては，F 値で 92.7%を示しており，本システムの分類手法の有効性を確認した．

なお，日本語文作成支援の代表的な方法として，用例の検索以外に，文の自動添削が挙げられる．文添削には，文の作成者

が気付いていない誤りを校正できるという利点がある．そのような研究として，ルールベースによる文添削の自動化が行われている [4–7]．しかし，文添削は，自分が書きたい内容をある程度表現できる人を対象としており，そもそも，書きたい内容を文として表現することが難しい人にとっては，十分活用できないという問題がある．

2. 関 連 研 究

日本語用例の検索に関する研究として，共起語や用例文を提示するシステムが提案されている [1,2]．

吉橋らは，日本語文作成支援システム「なつめ」を提案している [1]．このシステムでは，名詞を入力として，その名詞と共起する動詞を，名詞に後続する格助詞によって分類し，提示する．さらに，共起する動詞に対する用例文を提示できる．

深田は，日本語用例・コロケーション抽出システム「茶漉」を提案している [2]．ユーザはキーワード，キーワードに対する共起語，及び，共起語との距離の 3 項目を入力する．システムは検索結果として用例文と共起語の情報を提示する．

これらのシステムでは，キーワードの共起語に注目している．しかし,「なつめ」のように共起語や関連する用例文を提示するだけでは，検索結果にユーザにとって不必要な文も提示されてしまう．また,「茶漉」では，ユーザが共起語との距離を適切に入力することは難しいという問題がある．

3. 日本語文用例検索システム EscortJacs

3.1 システムの概要

本節では，用例検索システム EscortJacs について述べる．EscortJacs は留学生の日本語文作成支援を目的としたシステムである．留学生が作成した日本語文には，助詞の誤りが多いことが知られている [3]．そこで，本システムでは，日本語キーワード系列を入力として受け取り，キーワード同士を接続する適切な助詞を明らかにするような用例文を提示することにより，留学生に対する作文支援を行う．

以下の 4 文はいずれも,「医師」「話」「聞いた」というキーワードを含んでいる．

(3-a) 医師に 治療の 話を 聞いた。
(3-b) 医師は 患者から 話を 聞いた。
(3-c) 親と 一緒に、 医師の 話を 聞いた。
(3-d) 会場に 医師を 招いて、 医療の 話を 聞いた。

これらの文のうち，文 (3-a)，(3-b)，(3-c) は，用例文として適している．一方，文 (3-d) は，用例文として提示することは不適切である．(3-a),(3-b),(3-c) のようにキーワードをつなぐ助詞が存在する場合，キーワード間に直接的な関係が存在しているが，(3-d) のような場合は，直接的な意味関係が存在していない．

本研究では，キーワードを含む文節間に存在する直接的な意味の関係性を, 係り受け構造を考慮することによって捉える．係り受け構造とは文節間の係り受け関係の集合で示される．各用例文の係り受け構造を図 1 に示す．例えば，文 (3-a) は，文節「医師に」が文節「聞いた」に係り，文節「話を」が文節「聞いた」に係る．しかし，文 (3-d) は，キーワードを含む文節間に直接的な係り受け関係が存在しない．このように，キーワード間に直接的な係り受け関係が存在しない文の提示を避けることにより，キーワード同士を接続する適切な助詞を明らかにするような用例文のみを提示する．

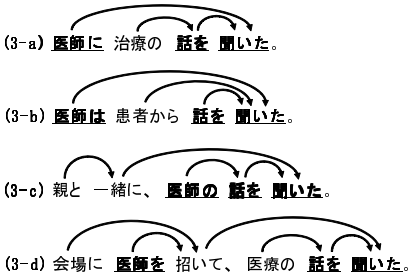

図 1 文の係り受け構造

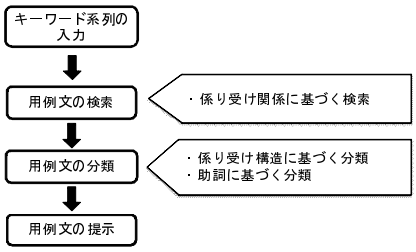

図 2 システムの構成

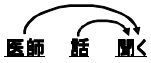

図 3 「医師 話 聞く」の係り受け構造パターン

本システムの構成を図 2 に示す．本システムは，検索部，および，分類部から構成される．検索部では，キーワード系列を受け取り，それらのキーワードを含む文節間に直接的な係り受け関係が存在する文を出力する．分類部では，検索部で得られた文を受け取り，それらの文を係り受け構造，及び，キーワードに後続する付属語に基づいて分類し，結果を出力する．システムは，分類された結果を提示する．

3.2 用例文の検索

3.2.1 係り受け関係に基づいた検索

検索処理では，入力されたキーワード系列を受け取り，キーワードを含む文節間に直接的な係り受け関係が存在する文を出力する．具体的には，まず，入力された各キーワードを含む文集合をそれぞれ得る．次に，それらの和集合をとることで，全てのキーワードを含む文集合を獲得する．獲得された文集合中の各文に対し，キーワードが出現する文節位置と文の係り受け構造を取得し，係り受け構造パターン生成アルゴリズムにより，係り受け構造パターン生成の成否を判定する．係り受け構造パターンは，キーワードを含む文節間の係り受け構造を表現するものである．例えば，文 (3-a) の,「医師」「話」「聞いた」を含む文節に対する，係り受け構造パターンは図 3 のように表現される．このアルゴリズムでは，キーワード間に直接的な係り受け関係が存在しない文には係り受け構造パターンが生成されないため，パターン生成の成否により，キーワード間に直接的な

係り受け関係が存在しているか否かを判定できる．

3.2.2 係り受け構造同定アルゴリズム

係り受け関係に基づいた検索は，依存関係に基づく英文検索手法 [8] を日本語に適用することで実現した．入力として，

クエリ $q_1 \cdots q_m (q_i (1 \leqq i \leqq m)$ はキーワード)

文 $s = c_1 \cdots c_n (c_j (1 \leqq i \leqq n)$ は文節)

文の係り受け関係の集合 $D = \{(j, k) | c_j$が c_kに係る $\}$

を受け取り，係り受け構造パターンの集合を出力する．係り受け構造パターンは 3 項組 $d = (h, L, F)$ である．h は単語位置であり，これを d の主辞と呼ぶ．L は係り受け構造パターンのリストであり，F は係り受け構造パターンに含まれている付属語のリストである．文節は 1 個の自立語と 0 個以上の付属語から構成される言語単位である．係り受け構造パターンは，クエリ $q_1 \cdots q_m$ に対して，以下の操作をボトムアップに適用することにより生成する．

初期化

各 $q_i (1 \leqq i \leqq m)$， $c_j (1 \leqq j \leqq n)$ に対して，q_i が文節 c_j に含まれ，その付属語 f_j が存在するならば，q_i に対する係り受け構造パターンとして (j, ϵ, f_j) を生成する．付属語が存在しないならば、q_i に対する係り受け構造パターンとして (j, ϵ, ϵ) を生成する．

結合操作

$d = (h, L, F)$，及び，$d' = (h', L', F')$ をそれぞれ，$q_i \cdots q_j$，及び，$q_{j+1} \cdots q_k (k \leqq m)$ に対する係り受け構造パターンとする．このとき $(h, h') \in D$ ならば，$q_i \cdots q_j q_{j+1} \cdots q_k$ に対する係り受け構造パターン (h', dL', FF') を生成する．

結合操作を繰り返し適用することにより，各キーワードを含む文節が別のキーワードを含む文節に直接係るようなすべての係り受け構造パターンを生成できる．

3.3 用例文の分類

3.3.1 係り受け構造と助詞・助動詞に基づく分類

本システムは，キーワード同士を接続する適切な助詞を明らかにするような用例文を提示することを目的としている．そのため，表現の違いや助詞により文を分類することにより，ユーザは，助詞の異なる使用法を区別して用例文を調べることができる．そこで，本システムでは用例文を係り受け構造に基づいて分類し，さらに，キーワードに後続する付属語ごとに分類して提示する．まず，係り受け構造の種類により，文を分類する．例えば，文 (3-a)，(3-b) は，図 1 に示される通り,「医師」,「話」を含む文節が，共に「聞いた」を含む文節に係る．一方，文 (3-c) は,「医師」を含む文節が「話」を含む文節に係り,「話」を含む文節が「聞いた」を含む文節に係る．そのため，文 (3-a) と (3-b) は同じクラスに分類され，文 (3-c) は異なるクラスに分類される．

また，文 (3-a) は「医師」が「聞く」の目的語であるのに対し，文 (3-b) では「医師」が「聞く」の主語である．このように，キーワード間の係り受け構造が同じであっても，係り先のキーワードとの関係が異なる場合がある．日本語では，付属語によって文節間の関係が決まるため，キーワードに後続する付属語に着目して分類することにより，ユーザは付属語の使用法を区別して用例文を調べることができる．そこで，本システムでは，係り受け構造によって同一クラスに分類された文に対して，キーワードに後続する付属語に着目した細分類を行う．例えば，文 (3-a)，(3-b) は,「医師」に後続する助詞の種類によって，異なるクラスに分類される．

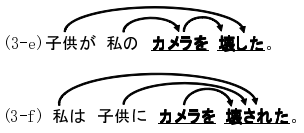

図 4 「カメラ」「壊す」を含む文に対する係り受け構造

さらに，文が受動態か能動態かを区別するため，入力末尾のキーワードに助動詞「れる・られる」が後続する場合，別のクラスに分類する．キーワード「カメラ」「壊す」に対して，以下の文が検索された場合を考える．

(3-e) 子供が 私の カメラを 壊した。

(3-f) 私は 子供に カメラを 壊された。

上記の文に対する係り受け構造を図 4 に示す．文 (3-e)，(3-f) は，共に「カメラ」を含む文節が「壊す」を含む文節に係り，キーワード「カメラ」に後続する付属語は「を」である．しかし，文 (3-f) には入力末尾のキーワード「壊す」に後続する付属語に「れる」が含まれているため，これらの 2 文は別のクラスに分類する．

3.3.2 分類アルゴリズム

係り受け構造パターンが同一の文をまとめることにより文を分類する．ただし，同一の係り受け構造パターン中の単語の出現位置については考慮しない．具体的には，以下に定義する同値関係 $\equiv$ が成り立つとき，同一の係り受け構造パターンとみなす．なお，$passive(f)$ は，付属語 f が「れる」，または,「られる」であれば 1 を返し，それ以外のとき 0 を返す関数である．また，l_i は L の i 番目の要素 (係り受け構造パターン) を表し，f_i は F の i 番目の要素（付属語の種類）を表す．

係り受け構造パターンの同一性

$d = (h, L, F)$，および $d' = (h', L', F')$ を係り受け構造パターンとする．条件 (1)〜(5) を全て満たすとき，$d \equiv d'$ と定義する．

(1) $|L| = |L'|$

(2) $|F| = |F'|$

(3) $\forall i (1 \leqq i \leqq |L|), l_i \equiv l'_i$

(4) $\forall i (1 \leqq i \leqq |F| - 1), f_i = f'_i$

(5) $passive(f_{|F|}) = passive(f_{|F'|})$

このように，依存構造パターン同定アルゴリズムにより，キーワード間の係り受け関係，及び，キーワードに後続する付属語に従って文を分類することができる．

4. EscortJacs の実装

4.1 システムの実装

日本語用例検索システム EscortJacs を Perl を用いて実装した．検索対象として，読売新聞記事データ'98，'99 年度の 2,344,625 文を用いた．これらの文には，形態素情報，係り受け構造の情報が付与されている．形態素解析器として MeCab [9]

図 5　EscortJacs の入力画面

図 6　EscortJacs の検索結果

を，構文解析器として CaboCha [10] を用いた．

本システムは Web 上で動作する．システムの入力画面を図 5 に示す．画面中央部は検索窓であり，ユーザはこの検索窓に各キーワードをスペースで区切って入力し，検索する．システムの検索結果を図 6 に示す．検索結果は用例文の表示部分と分類結果の表示部分で構成される．分類結果の表示部分では，係り受け構造，及び，付属語によって分類されたクラスを提示する．また，用例文の表示部分では，上記のクラスごとに分類された用例文を提示している．

4.2　システムの動作例

図 6 は，キーワード系列「家」「外」「出る」に対する検索結果を示している．左側の分類結果の表示部分では，まず，そのクラスに分類される用例文の数が多い順に，クラスを提示する．パターン 1 では，「家」を含む文節が「外」を含む文節に係り，「外」を含む文節が「出る」を含む文節に係るような係り受け構造を示している．さらに，各係り受け構造に対して，キーワードに後続する付属語に基づいて分類し，キーワード「家」「外」「出る」を含む表現を，文数が多いものから順に表示する．パターン 1 では，「家の外に出る」という表現を含む文が 8 文，「家の外へ出る」という表現を含む文が 2 文であるため，「家の外に出る」，「家の外へ出る」の順に表示する．また，各分類に含まれる文の数を表示することにより，ユーザは各表現の使用頻度を確認できる．

右側の用例文の表示部分では，用例文を分類結果と同じ順序で提示する．ある分類に含まれる用例文が 3 文以上ある場合は，ユーザは「もっと見る…」というリンクから，そのクラスに該当する全ての用例文を確認できる．

5.　評 価 実 験

本システムの有効性を確認するために，本システムの特徴である係り受け関係に基づく検索，及び，係り受け構造と付属語に基づく分類性能について評価実験を行った．

5.1　実 験 概 要

用例検索では，ユーザにとって適切な用例文を提示することが求められる．そのため，検索手法の性能を評価するためには，提示された用例文のうち，どれだけの文がユーザにとって適切であるかを判定する必要がある．そこで，作業者が作成した正解データを用いて評価実験を行った．

正解データの作成は，大学に在学中の留学生 5 名が以下の手順で実施した．

（1）　日本語文の助詞の穴埋め問題が提示される．

（2）　提示された問題を解くために，用例検索に用いるクエリを作成する．なお，クエリはスペースで区切られたキーワード列からなる．

（3）　作成したクエリ中のキーワードをすべて含む文を検索し，提示する．

（4）　提示された用例文のクエリに対する適切性を判定する．

手順 (1) で提示する問題は，日本語の誤用に関する論文 [3]，書籍 [11]，及び，日本語能力試験問題集 [12] を参考に作成した．

手順 (3) において，作成したクエリに対して大量の文が提示される場合があり，その全てを判定することは，作業者にとって多大な労力を伴う．そのため，作業者に提示する用例文として 30 文程度を抽出した．用例文の抽出手順は以下の通りである．まず，各クエリに対して検索された用例文を，依存構造パターンとキーワードに後続する付属語によって分類する．ただし，クエリ中の各キーワード間に直接的な係り受け関係が存在しない文については，それらを 1 つのクラスとして分類する．次に，各クラスに属する文数の割合に従い，検索された用例文のうちから合計 30 文程度になるように用例文を抽出する．この際，各クラスから少なくとも 1 文抽出する．検索された用例文の合計数が 30 文に満たない場合は，全ての用例文を提示する．

また，用例文を提示する際には，係り受け情報と分類の情報が手順 (4) における作業者の判定に影響を与えないようにするため，用例文のみを提示した．

作成した正解データの規模を表 1 に示す．5 人の作業者による検索回数は 67 回であった．67 回のうち，用例文が提示されたのは 53 回であった．すなわち，正解データに含まれるクエ

表 1　正解データの規模

クエリ数	53
キーワード検索による文数	1186
正解文数	454

表 2　検索性能の評価結果

	精度 (%)	再現率 (%)	F 値
ベースライン	39.2	100.0	56.3
提案手法	59.6	89.1	71.5

リ数は 53 個である．ただし，異なる作業者が同じクエリを用いて検索を行った場合は，別のクエリとして扱っている．これは，同じクエリに対しても作業者ごとに正解となる文が異なると考えられるためである．53 個のクエリに対し，1186 文が提示された．

5.2　検索性能の評価実験

係り受け構造に基づく検索の有効性を評価した．正解データに含まれる 53 個のクエリを用いて，係り受け構造に基づく検索によって得られた 708 文を評価対象とした．評価指標には精度と再現率，及び，それらの調和平均である F 値を用いた．各指標として，作業者ごとの精度と再現率，F 値の値を計算し，それらを平均した値を用いた．作業者ごとの精度，及び，再現率は以下の式で示される．

$$\text{精度} = \frac{\text{検索により提示された正解文数}}{\text{検索により提示された文数}} \tag{1}$$

$$\text{再現率} = \frac{\text{検索により提示された正解文数}}{\text{正解文数}} \tag{2}$$

実験結果を表 2 に示す．ベースライン手法によって提示された文は，正解データを作成する際に提示された文と同一であるため，再現率は 100%である．ベースラインと比較し，提案手法による検索の精度は 59.6%と向上している．また，F 値において，提案手法は 71.5%と高い値を示しており, 提案手法の有効性を確認した．

5.3　分類性能の評価実験

本システムの分類手法の有効性を評価した．分類実験において，係り受け関係が存在する文が検索されなかったクエリに対しては，分類を行うことができない．そのため，53 個のクエリのうち，そのような 4 個のクエリを除いた 49 個のクエリを用いて用例文を検索した．各クエリに対して得られた用例文に対して，提案する分類手法に基づいて分類を行い，分類性能を評価した．評価には，加藤らの方法 [13] を利用した．以下で，この評価方法について説明する．

用例文の分類により生成されたクラスを $d_1, \cdots, d_l$ ，クラス $d_i (1 \leqq i \leqq l)$ に含まれる日本語用例文の集合を S_{d_i} とする．このとき，正解文のうち提案手法により検索される文集合 C に関して，C と完全に一致する文集合 S_d^* が，検索結果 $\{S_{d_i}, \cdots, S_{d_l}\}$ の中に存在するのが理想的な分類である．そこで，C と一致する文集合が，検索結果 $\{S_{d_i}, \cdots, S_{d_l}\}$ に存在するか否かにより分類を評価する．ただし，C と完全に一致する S_{d_i} が存在しない場合についても，その一致の度合いを評価するために，精度，再現率，及び，F 値を用いる．具体的な評価方法は以下の通りである．

表 3　分類性能の評価結果

	精度 (%)	再現率 (%)	F 値
ベースライン 1	64.8	100.0	78.6
ベースライン 2	100.0	29.3	45.3
提案手法	94.3	91.6	92.9

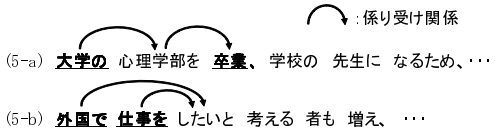

図 7　間接的な係り受け関係が成立する文

（1） 各 S_{d_i} について，精度 P_{d_i} と再現率 R_{d_i} により C との一致の度合いを評価する．

（2） 精度と再現率の F 値が最大である （すなわち， C と最も一致している）S_{d_i} を S_d^* とする．その精度と再現率を 1 つのクエリに対する分類の精度と再現率とする．分類手法の精度と再現率は各作業者ごとの精度と再現率の値を計算し，それらを平均した値を用いた．この評価方法では，C と完全に一致する S_{d_i} が存在するならば，分類の精度と再現率は共に 100% である．S_{d_i} の精度 P_{d_i} と再現率 R_{d_i} は次のように定義する．

$$\text{精度 } P_{d_i} = \frac{|S_{d_i} \cap C|}{|S_{d_i}|}, \quad \text{再現率 } R_{d_i} = \frac{|S_{d_i} \cap C|}{|C|} \tag{3}$$

比較のため，2 種類のベースラインを設定した．ベースライン 1 は検索結果全体を 1 つのクラスに分類する手法，ベースライン 2 は全ての用例文を別のクラスに分類する手法である．

分類の評価結果を表 3 に示す．ベースライン 1 は 1 つのクラスに全ての正解文を含んでいるため，再現率が 100%となる．ベースライン 2 はいずれかのクラスに必ず 1 つの正解文を含む分類手法のため，精度が 100%となる．2 種類のベースラインに比べ，提案手法は F 値で 92.9%と高い値を示しており，本方式の分類手法の有効性を確認した．

5.4　考　　察

5.4.1　検索性能に関する考察

実験において，本手法によって検索されなかった正解文が 51 文存在した．それらの正解文が検索されなかった原因について分析を行った．

検索されなかった正解 51 文のうち 13 文には，入力したキーワードを含む文節間に，1 文節を介した間接的な係り受け関係が存在した．そのような正解文の例を図 7 に示す．文 (5-a), (5-b) はキーワードを含む文節間に直接的な係り受け関係が存在しないため，本手法では検索されない．

しかし，文 (5-a)，(5-b) はキーワードを含む文節間に，1 文節を介した係り受け関係が存在する．直接的な係り受け関係以外に間接的に成立している係り受け関係を考慮することにより，

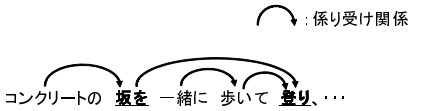

図 8　正しい係り受け構造

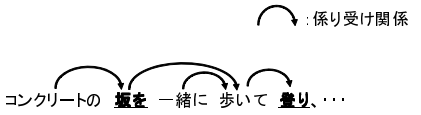

図 9　誤って解析された係り受け構造

これらの文も検索可能となると考えられる．しかし，単純に間接的な係り受け関係が成立している文を検索するだけでは，不適切な用例も検索されることになり，検索の精度が下がる可能性がある．

また，51 文のうち 10 文において係り受け解析誤りが存在した．そのような例として，以下の文があった．

- **コンクリートの坂を一緒に歩いて登り、ケーブルテレビアンテナ基地フェンス近くの山道で首を絞めた、などと供述。**

この文に対する正しい係り受け構造を図 8 に，誤って解析された係り受け構造を図 9 に示す．本来，文節「坂を」と「登り」の間に直接的な係り受け関係が存在するものの，解析結果では，「坂を」と「登り」の間に直接的な係り受け関係が成立せず，本手法によって検索されなかった．

5.4.2　分類性能に関する考察

本手法によって正しいクラスに分類されなかった正解文が 38 文存在し，その原因を調査した．

38 文のうち，12 文は，問題文の文脈から，複数の助詞が正解として考えられる場合であった．例えば，

- **彼女は科学にはあまり興味 __ なさそうです。**

という問題では，作業者の 1 人が「興味」「ない」というクエリで検索し，以下のような文が提示された．

(5-c)　ごみ焼却炉の周辺に住む人は恐れているが、生徒はあまり興味がない。

(5-d)　映画に興味はなかったが、そこで八ミリ映画と出合った。

問題文の文脈から判断すると,「興味」と「ない」を接続する助詞として,「が」と「は」はいずれも適切である．作業者は，この 2 文をともに正解と判定した．しかし，今回の評価手法では，正解数が最大となるクラスを正解の分類としているため，文 (5-d) を不正解としている．すなわち，接続する助詞として適したものが複数存在する場合も考慮する必要があるといえる．

また，3 文は助動詞「れる・られる」を含む文であった．本手法では，能動態の文と受動態の文を別のクラスに分類するため，助動詞「れる・られる」に着目した分類を行っている．しかし，助動詞「れる・られる」は受動態を表すだけではなく，自発，可能，尊敬の意味を持つ場合があるため，単純に「れる・られる」を含む文が受動態の文であるとはいえない．例えば，クエリ「質問　答える」に対する正解文

(5-e)　生まれる子供の 質問に答えられたら、と思い聞き続けた。

の助動詞「れる」は可能の意味を持つ．しかし，本手法では，単純に「れる・られる」を含むか否かによって分類しているため，文 (5-e) を正解とは異なるクラスに分類している．すなわち，助動詞「れる・られる」の用法を考慮する必要がある．そのために，受身の用法で使われている「れる・られる」の共起語の情報を，コーパス等から獲得し,「れる・られる」の分類に利用することで，分類誤りを減らすことができると考えられる．

6. おわりに

本論文では，日本語文作成支援のための用例検索システムを提案した．本システムでは，日本語キーワード系列を入力とし，キーワードを含む文節間に直接的な係り受け関係が存在する文のみを用例文として提示する．係り受け関係を考慮することにより，ユーザにとって有用な用例文を検索することが可能となる．また，用例文の提示では，キーワード間の係り受け構造，及び，キーワードに後続する付属語に基づいて用例文を分類することにより，ユーザが適切な用例文を容易に発見可能な提示方法を実現した．

本システムにおける検索と分類の有効性を確認するために，2 種類の評価実験を行った．検索性能の評価実験において，提案手法の検索精度は単純なキーワード検索と比べ，20.4 ポイント向上し，係り受け関係の利用が，検索精度の向上に寄与することを確認した．また，分類性能の評価実験においては，F 値で 92.7%を示し，係り受け構造，及び，キーワードに後続する付属語に基づく分類手法の有効性を確認した．

文　献

[1] 吉橋, 曹紅, 仁科：“作文支援システム「なつめ」における共起表現表示機能と評価”, 日本語教育方法研究会, **14**, 1, pp. 44–45 (2007).
[2] 深田：“日本語用例・コロケーション情報抽出システム「茶漉」”, 日本語科学, **22**, pp. 161–172 (2007).
[3] 細川：“留学生日本語作文における格関係表示の誤用について”, 早稲田大学日本語研究教育センター紀要, pp. 70–89 (1993).
[4] 杉野, 佐藤, 絹川：“外国人の初級日本語文における振り仮名の誤り検出”, 情報科学技術フォーラム, **8**, pp. 591–592 (2009).
[5] 佐藤, 杉野, 絹川：“外国人の初級日本語文における振り仮名の誤り訂正”, 情報科学技術フォーラム, **8**, pp. 593–594 (2009).
[6] 南保, 乙武, 荒木：“文節内の特徴を用いた日本語助詞誤りの自動検出・校正”, 情報処理研究報告, **2007**, 94, pp. 107–112 (2007).
[7] 綿貫：“統計的言語モデルを用いた日本語活用誤りチェッカ”, Master's thesis, 筑波大学 (2005).
[8] 江川, 加藤, 松原：“ラベル付き依存関係に基づく英文用例検索システム”, 言語処理学会第 13 回年次大会発表論文集, **13**, pp. 294–297 (2007).
[9] 工藤, 山本, 松本：“Conditional random fields を用いた日本語形態素解析”, 情報処理研究報告, **2004**, 47, pp. 89–96 (2004).
[10] 工藤, 松本：“チャンキングの段階適用による日本語係り受け解析”, 情報処理学会論文誌, **43**, 6, pp. 1834–1842 (2002).
[11] 鈴木：“教師用日本語教育ハンドブックシリーズ 3 文法 I”, 凡人社 (1992).
[12] 国書刊行会：“日本語能力試験直前対策 文法 3 級” (1998).
[13] 加藤, 江川, 松原, 稲垣：“依存構造に基づく用例文検索手法とその評価”, 電子情報通信学会論文誌, **J92-D**, 3, pp. 417–427 (2009).