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.

# 大規模コーパスを用いた日本語機能表現呼応可能性規則の作成および日本語学習支援での利用

黒川　和也†　　宇津呂武仁†

† 豊橋技術科学大学 工学部 情報工学系
〒 441–8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1
E-mail: †{kazuya,utsuro}@cl.ics.tut.ac.jp

あらまし　本論文では，自然言語文における正用と誤用の区別を規則化するという課題の一例として，接続助詞と助動詞が一文中で呼応する複文に焦点を当て，周囲の自立語によらず，どのような場合でも，接続助詞と助動詞の呼応が許されない場合について調査を行った．大規模コーパスを用いた呼応事例収集と机上実験を行い，二万組以上の接続助詞と助動詞の組合せについて呼応可能性を調査した結果，約四分の三の組合せが呼応可能であることが分かった．さらに，接続助詞あるいは助動詞の練習問題の誤り選択肢を自動生成する過程において，作成された接続助詞・助動詞の呼応可能性規則の評価を行い，5∼8 割程度の精度で適切な誤り選択肢が自動生成できた．
キーワード　コーパス，日本語教育，日本語学習支援，機能表現，文法練習問題，誤用

# Judging Well-formedness of Collocation of Japanese Functional Expressions based on Large Scale Corpora and Their Use in Japanese Learning Assistance

**Kazuya KUROKAWA† and Takehito UTSURO†**

**† Dpt. of Information and Computer Sciences, Faculty of Eng., Toyohashi Univ. of Tech.**
**Tenpaku-cho, Toyohashi, Aichi 441–8580, Japan**
**E-mail: †{kazuya,utsuro}@cl.ics.tut.ac.jp**

**Abstract**　This paper reports results on examining well-formedness of collocation of Japanese conjunctive particles and auxiliary verbs within a sentence. Judgment on well-formedness is based on the results of collecting collocation from large scale corpora and also on manual examination. About 75 % of more than 20,000 pairs of conjunctive particles and auxiliary verbs are judged as well-formed. Results of judging well-formedness are evaluated in the process of automatic generation of exercises of conjunctive particles and auxiliary verbs, achieving 50∼80 % accuracies.
**Key words**　Corpus, Japanese Language Education, Japanese Learning Assistance, Functional Expressions, Exercises for Grammar Learning, Language Use Errors

## 1. はじめに

言語表現の用法上の規則性を記述し，自然言語文の正用と誤用を明確に区別するという問題は，自然言語処理や語学教育など言語に関する研究分野において，最も基本的で重要な課題の一つである．例えば，文生成システムや機械翻訳システムにおいて，自然な文あるいは訳文を生成するためには，自然言語文の正用と誤用の明確な区別が不可欠である．また，作文推敲支援システムにおいては，文法上の誤用を検出して適切な訂正を行う機能が必要である．一方，語学学習支援システムなどにおいても，学習者の誤用を検出して適切な訂正を行う機能が必要であり，また，語学学習用練習問題のオーサリングシステムにおいては，正用と誤用を的確に区別して，適切な選択肢の候補を生成する機能が必要である．

このような観点で，我々は，日本語学習における文法習得に焦点を当て，日本語文法習得において重要な役割を持つ機能表現の正用と誤用をいかにして区別するかという問題について研究を行ってきた [1]．そして，日本語機能表現の正用と誤用の違いの規則性を利用して，日本語能力試験 [3] において出題の対象となる接続・修飾・文末などの機能表現の練習問題を，日本語コーパスから半自動的に生成する方法について研究を行ってきた．この研究では，機能表現の用法上の正誤を左右する要

表 1: 意味分類および統語的接続形式による日本語機能表現の分類 (機能表現例)

| | | |
|---|---|---|
| 非文末型 | | |
| 格助詞タイプ | 体言接続 | 資格・立場 (をもって，でもって，etc.), 対象・関連 (について，に関して，etc.)，仕手・仲介・手段・根拠・原因 (によって，を通じて，etc.)，時・場所・状況 (に際し，につけ，etc.)，起点・終点・範囲 (をはじめ，にかけて，etc.)，基準・境界 (をもって，etc.)，割合 (に対して，etc.)，対応 (によると，etc.)，同格 (ところの，etc.)，<br>全 9 大分類，30 細分類，63 種類 |
| | 用言接続 | 時・場所・状況，同格，全 2 大分類，8 細分類，16 種類 |
| 係助詞タイプ | 体言接続 | 定義 (とは，というのは，といったら，etc.)，主題化 (といえば，といって，etc.)，全 2 大分類，13 細分類，30 種類 |
| | 用言接続 | 主題化，全 1 大分類，12 細分類，25 種類 |
| 副助詞タイプ | 体言接続 | 強調 (という，といった，etc.)，限定・非限定 (に限って，にかぎり,etc.)，添加 (に限らず，のみならず，etc.)，除外 (をよそに)，不適合 (にもなく，etc.)，不明確 (ともなく，etc.)，全 6 大分類，17 細分類，25 種類 |
| | 用言接続 | 強調，全 1 大分類，8 細分類，6 種類 |
| 接続助詞タイプ | 用言接続 | 1:同時性 (や否や，が早いが，そばから，etc.)，2:継起 (あげく，ところ，etc.)，3:相関 (に従い，につれて，etc.)，4:順接-仮定 (かぎりは，ようものなら，etc.)，5:順接-確定 (てみると，てみたら，etc.)，6:因果関係 (からには，おかげで，etc.), 7:逆接-仮定 (としても，にしても，にしたって，etc.)，9:逆接-確定 (からといって，にもかかわらず，etc.)，9:対比 (のに対して，と同時に，etc.)，10:反復・並立・添加 (まま，ばかりか，etc.)，全 10 大分類，41 細分類，129 種類 |
| 文末型 | | |
| 終助詞タイプ | 体言接続 | 感動・詠嘆・驚異，疑問・問いかけ・確認，伝聞，回想，断定・強調，全 5 大分類，18 細分類，21 種類 |
| | 用言接続 | 感動・詠嘆・驚異 (とは，ものだ，etc.)，疑問・問いかけ・確認 (かい，ではないか，etc.)，反語・反駁・否定・非難・後悔 (ものか，くせに，etc.) 願望・勧誘 (ないかな，ようか，etc.)，伝聞 (とのこと，ということ，etc.)，回想 (たっけ，etc.)，適当 (ものか)，理由・根拠 (ことだし，etc.)，断定・強調 (というものだ)，全 9 大分類，35 細分類，46 種類 |
| 助動詞タイプ | 体言接続 | 義務・当然・当為・必然・必要・勧告・主張，当為等の否定・不必要，推量・推測・推定，限定，程度，経験・回想・習慣，伝聞，全 7 大分類，12 細分類，24 種類 |
| | 用言接続 | 1:禁止 (てはいけない，etc.)，2:義務・当然・当為・必然・必要・勧告・主張 (なければならない，はずだ，etc.)，3:当為等の否定・不必要 (ものではない，とは限らない，etc.)，4:可能・不可能 (ことができる，わけにはいかない，etc.)，5:許容・許可 (てもいい，てもかまわない，etc.)，6:意志・決定 (ようとする，つもりだ，etc.)，7:自然成立・自発・強制・肯定的意向の強調 (ことになる，ざるを得ない，etc.)，8:推量・推測・推定 (かもしれない，はずだ，etc.)，9:適当・願望・提案・勧誘・勧告 (ばいい，てほしい，etc.)，10:要求・依頼 (てくれ，てください，etc.)，11:限定 (よりほかない，までだ，etc.)，12:程度 (に過ぎない，だけのことはある，etc.)，13:経験・回想・習慣 (ことがある，ものだ，etc.)，14:伝聞 (とのことだ，etc.)，15:行為の授受 (てやる，てもらう，etc.)，16:アスペクト (ずにいる，てしまう，ずにしまう，etc.)，全 16 分類，74 細分類，177 種類 |
| 合計 | | 助詞タイプ: 36 大分類・136 細分類・273 種類，助動詞タイプ:16 大分類・74 細分類・177 種類 |

因として，文中の他の機能表現だけでなく，周囲の自立語による影響もかなり大きいことが分かった．このため，ある機能表現について，周囲の自立語の影響を無視して，あらゆる可能な (文内) 文脈における正用・誤用の区別を規則化することは容易ではなかった．ただし，複文などのように，機能表現の周囲の (文内) 文脈が複雑な場合には，一文中の二つ以上の機能表現の間の呼応に関する制約などの条件が厳しくなり，周囲の自立語とは独立に正用・誤用の区別が規則化できる事例もあった．

そこで，本論文では，機能表現のうち，特に，接続助詞と助動詞が一文中で呼応する複文に焦点を当て，周囲の自立語によらず，どのような場合でも，接続助詞と助動詞の呼応が許されない場合について調査を行った．ここで，接続助詞と助動詞が一文中で呼応可能であるか否かという現象は，接続助詞あるいは助動詞単独の用法上の制約のごく一部に過ぎないが，文内の自立語の影響を受けずに正用・誤用の規則性を記述できるという大きな利点がある．さらに，接続助詞と助動詞の呼応可能性を分析することにより，接続助詞あるいは助動詞単独の用法上の制約の手がかりが得られる可能性がある．

調査の具体的な方法として，まず，新聞記事コーパス十数年分に対して接続助詞と助動詞の呼応の事例を有無を調査し，呼応事例が収集できなかった接続助詞と助動詞の組合せについて，呼応が可能かどうかの机上実験を網羅的に行った．その結果，二万組以上の接続助詞と助動詞の組合せについての呼応可能性規則が得られ，そのうち約四分の三の組合せが呼応可能であることが分かった．また，接続助詞あるいは助動詞の練習問題の誤り選択肢を自動生成する過程において，作成された接続助詞・助動詞の呼応可能性規則の評価を行い，5∼8 割程度の精度で適切な誤り選択肢が自動生成できた．本論文では，これらの過程について報告し，さらに，接続助詞・助動詞呼応の誤用の自動訂正などの発展的課題において，作成された接続助詞・助動詞の呼応可能性規則を利用する方法の考え方を述べる．

## 2. 日本語機能表現の分類体系

本論文では，接続助詞と助動詞の呼応関係を規定する際の参考情報として，接続助詞および助動詞の意味分類を補助的に用いている．ここで，これらの機能表現の意味分類は，日本語機能表現学習教材作成支援の研究 [1] において我々が用いているものである .[1] においては，意味および統語的接続形式の両方を考慮して，日本語機能表現を表 1 のように分類している．本節では，この日本語機能表現の分類体系の概要を述べる．

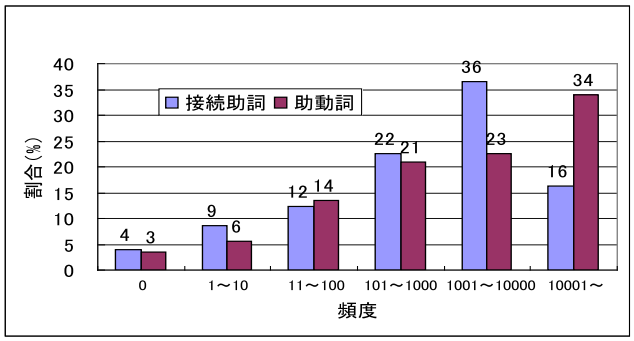


図 1: 接続助詞・助動詞単独での頻度分布

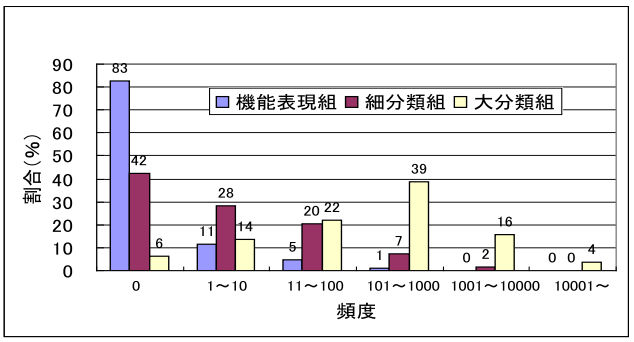


図 2: 接続助詞・助動詞の呼応の頻度分布

### 2.1 階層的意味分類体系

まず，機能表現の階層的意味分類としては，[2] の複合辞の階層的分類を参考にして分類を行う．表 1 に示すように，[2] の階層的意味分類においては，機能表現は，助詞と同様の働きをするもの (36 大分類・136 細分類・273 種類) と，助動詞の働きをするもの (16 大分類・74 細分類・177 種類) とに大きく分類され，助詞と同様の働きをするものの中で，格助詞タイプや，係助詞タイプのように文法機能によって細かく分類される．さらに，それぞれのタイプの中で，機能表現の意味による詳細な分類がなされている (表中で，接続助詞および助動詞の意味分類につけられている番号は，表 2 で参照するためのもの)．

### 2.2 統語的接続パターンによる分類

統語的接続パターンを用いた機能表現の分類においては，大きく，体言に接続するか，用言に接続するか，によって機能表現を分類する．さらに，それらの統語的接続パターンごとに，格助詞・係助詞・副助詞・接続助詞タイプのように，文の途中に現れるもの (以下，「非文末型」とよぶ) と，終助詞・助動詞タイプのように，(特に，練習問題の問題文においては) 文末に現れることが多いもの (以下，「文末型」とよぶ) とに分類する．

## 3. コーパスからの接続助詞・助動詞呼応事例の収集

大規模日本語コーパスに対して，接続助詞と助動詞が呼応関係にある文を収集した．ここで，文中の接続助詞と助動詞が呼応関係にあるか否かについては，既存の係り受け解析ツールを用いて判定した．大規模日本語コーパスとしては，読売新聞 14 年分 (1988~2001 年，約 3,000 万文，2.1Gbyte) を用い，接続助詞および助動詞としては，表 1 に示したもの (接続助詞 129 種類，助動詞 177 種類) を収集の対象とした．具体的には，個々の文に対して逐次的に以下の処理を行い，条件を満たす文を収集する，という手順で収集作業を行った．

( 1 ) 一文に対して，接続助詞および助動詞の個々の表現との文字列照合を行い，文字列レベルで，接続助詞もしくは助動詞を含むと判定される場合には，次の処理に進む．

( 2 ) 一文に対して，形態素解析(注1) を行い，形態素列レベルで，接続助詞もしくは助動詞を含むか否かの照合を行い，接続助詞もしくは助動詞を含むと判定される場合には，その接続助詞もしくは助動詞の頻度カウントを一つ増やしてその文を記録し，次の処理に進む．

( 3 ) 形態素列レベルの判定として，一文中に，接続助詞および助動詞がこの順で含まれるか否かの判定を行う．接続助詞および助動詞がこの順で含まれる場合には，次の処理に進む．

( 4 ) 一文に対して，係り受け解析(注2) を行い，接続助詞が助動詞に係る場合には，接続助詞と助動詞が呼応関係にあると判定し，その接続助詞と助動詞の呼応関係の頻度カウントを一つ増やしてその文を記録する．

この収集作業の結果得られた，接続助詞および助動詞単独での頻度分布を，接続助詞あるいは助動詞の表現レベルで求めたものを図 1 に示す．また，接続助詞と助動詞の呼応関係の頻度分布を表現レベル ($129 \times 177 = 22,833$ 組) で求めたもの，表 1 の意味分類の細分類レベルで求めたもの，および，大分類で求めたものを図 2 に示す(注3)．この結果から分かるように，新聞 14 年分のコーパスにおいて一度も出現しない接続助詞および助動詞はごくわずかである．しかし，実際に呼応関係になった接続助詞と助動詞の組合せは，組合せ全体の二割以下と少数であった．ただし，意味分類の細分類の組合せにまで広げると半数以上の組合せの呼応関係の事例がコーパス中に存在しており，意味分類の大分類では，9 割以上の組合せの呼応関係の事例がコーパス中に存在している．

## 4. 接続助詞・助動詞呼応可能性規則の作成

次に，前節で述べた全 22,833 組の接続助詞と助動詞の組合せについて，実際に呼応が可能かどうかの判定を人手で行い，その結果を，接続助詞・助動詞呼応可能性規則とした．ここで，呼応が可能かどうかの判定は，

| | |
|---|---|
| 呼応可能: | 当該接続助詞と当該助動詞の組合せを呼応させることができる例文が存在する |
| 呼応不可能: | どのような例文においても，当該接続助詞と当該助動詞の組合せを呼応させることは不可能 |

(注1): 日本語形態素解析システム茶筌 (http://chasen.aist-nara.ac.jp/)

(注2): 日本語係り受け解析システム CaboCha (http://cl.aist-nara.ac.jp/~taku-ku/software/cabocha/)

(注3): これらの頻度分布は，計算機による自動処理によって求めた結果であり，収集作業の過程で様々なエラーが混入している可能性がある．例えば，i) (2) において，接続助詞あるいは助動詞の前後に現れ得る形態素の条件まで含めて，厳密に形態素列レベルのパターンを網羅することまではできていないため，相対的に緩い検出となっており，接続助詞あるいは助動詞単独の頻度を大きめに見積もっている可能性がある．ii) (4) において，括弧表現を含む文の場合で，接続助詞と助動詞の間に括弧が含まれており，適切な呼応関係となっていない場合を収集している可能性がある．iii) 複数の意味分類に分類される接続助詞もしくは助動詞の場合，意味の類別が行われていないので，得られている頻度分布は，接続助詞もしくは助動詞の表現そのものに対する頻度分布である．このうち，i) および ii) については，今後，比較的容易に対処することが可能である．

表 2: 呼応可能な接続助詞・助動詞の組の割合— (呼応可能な接続助詞・助動詞組数/意味大分類中の接続助詞・助動詞組数) (%)

| 意味大分類番号 | 助動詞 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 合計 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 接続助詞 | 種類数 | 10 | 20 | 17 | 7 | 11 | 8 | 15 | 9 | 12 | 12 | 8 | 8 | 6 | 3 | 7 | 24 | 177 |
| 1 | 19 | 0 | 85 | 53 | 71 | 0 | 11 | 96 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 16 | 27 | 24 |
| 2 | 12 | 100 | 100 | 67 | 74 | 50 | 75 | 60 | 57 | 69 | 61 | 50 | 44 | 56 | 100 | 86 | 68 | 70 |
| 3 | 4 | 0 | 100 | 88 | 100 | 100 | 100 | 100 | 53 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 67 | 88 |
| 4 | 12 | 100 | 100 | 83 | 92 | 50 | 49 | 100 | 100 | 83 | 83 | 100 | 51 | 83 | 100 | 100 | 79 | 85 |
| 5 | 7 | 48 | 100 | 58 | 100 | 100 | 57 | 91 | 89 | 61 | 43 | 100 | 38 | 93 | 100 | 43 | 65 | 77 |
| 6 | 21 | 100 | 100 | 100 | 100 | 71 | 95 | 100 | 100 | 100 | 43 | 81 | 46 | 100 | 100 | 96 | 72 | 87 |
| 7 | 12 | 92 | 100 | 95 | 90 | 100 | 100 | 80 | 100 | 100 | 100 | 100 | 46 | 100 | 100 | 100 | 100 | 89 |
| 8 | 25 | 88 | 92 | 96 | 80 | 80 | 76 | 74 | 86 | 87 | 68 | 88 | 68 | 69 | 100 | 75 | 85 | 85 |
| 9 | 9 | 67 | 67 | 56 | 67 | 67 | 100 | 41 | 67 | 67 | 11 | 56 | 39 | 100 | 100 | 86 | 100 | 66 |
| 10 | 8 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 合計 | 129 | 74 | 94 | 83 | 85 | 65 | 72 | 71 | 75 | 74 | 58 | 72 | 48 | 73 | 100 | 77 | 68 | **74** |

のいずれであるかの判断を筆者らが行った．ただし，その際には，以下の手順にしたがった．

（1） 前節で述べた，接続助詞と助動詞の呼応関係の収集結果を参照して，収集の際に混入した誤りではなく，当該接続助詞と助動詞の正しい呼応の例が収集できている場合には，呼応可能であると判定する．

（2） その他の場合は，上記の判断を机上実験により行う(注4)．

以上の手順に基づいて，接続助詞・助動詞呼応可能性を判定した結果を集計したものを表 2 に示す．この表では，縦の欄に接続助詞の意味分類の大分類およびその大分類の接続助詞種類数，横の欄に助動詞の意味分類の大分類およびその大分類の助動詞種類数をそれぞれ示し，接続助詞および助動詞の意味分類の大分類の組合せの各々について，呼応可能な接続助詞と助動詞の組合せがどれだけあるかの割合 (当該意味分類中で呼応可能な接続助詞と助動詞の組合せ数/当該意味分類中の接続助詞と助動詞の全組合せ数) の一覧を示す．

新聞 14 年分のコーパスでは，実際に呼応関係になった接続助詞と助動詞の組合せは，組合せ全体の二割以下と少数であったにも関わらず，本節で作成した呼応可能性規則では，全組合せの四分の三近くが呼応可能と判定された．また，意味分類の大分類別に見ると，接続助詞では，時間的関係–同時性に分類される接続助詞との呼応可能性が極端に低かった．これは，呼応関係において，時間的関係–同時性に分類される接続助詞が文末表現に課す制約が厳しいためである．一方，反復・並立・添加に分類される接続助詞の場合は，文末表現に課す制約が少なく，どのような助動詞とも呼応可能であると判定された．助動詞の場合は，程度に分類される助動詞や要求・依頼に分類される助動詞の呼応可能性が低かった．一方，伝聞に分類される助動詞はどのような接続助詞とも呼応可能であると判定された．また，義務・当然・当為・必然・必要・勧告・主張に分類される助動詞の呼応可能性も高かった．

(注4): 前節で述べたように，実際に呼応関係になった接続助詞と助動詞の組合せは，組合せ全体の二割以下と少数であり，無条件に呼応可能であると判定できる場合は少ない．しかし，実際には，意味分類の細分類の組合せにまで広げると半数以上の組合せの呼応関係の事例がコーパス中に存在しており，これらの場合には，当該接続助詞と当該助動詞の組合せの呼応事例そのものがコーパス中に存在していなくても，同義接続助詞と同義助動詞の組合せの呼応事例が存在していることになる．したがって，全体の半数程度については，無条件に呼応可能であると判定できた．

## 5. 機能表現学習用練習問題自動生成における評価

接続助詞・助動詞について，以下の二種類の練習問題を自動生成する過程 [1] において，接続助詞と助動詞の呼応事例を正用例文として，前節で作成した呼応可能性規則の性能を評価した．

### 5.1 空欄補充問題

空欄補充問題とは，下記の問題例のように，空欄に適当な表現を選択肢から選択する問題である．

問題例 **1**

ライバルも好調な (　　)，油断するわけにはいかない．
1. だけに　　2. ところ

正用例文から空欄補充問題を作成する際には，正解選択肢 (この場合は選択肢 1) については，元々の正用例文の接続助詞 “だけに” (意味分類: 順接条件–因果関係) をそのまま用い，誤り選択肢については，助動詞 “わけにはいかない” (意味分類: 否定形による当為等・当為等の否定・不必要) との呼応が不可能な接続助詞 “ところ”(意味分類: 時間的関係–継起) を用いている．ここで，本節の評価では，誤り選択肢の自動生成の過程において，接続助詞と助動詞の呼応を含む正用例文に対して，4. で作成した接続助詞・助動詞呼応可能性規則を参照し，正用例文中の助動詞 (あるいは接続助詞) と呼応不可能とされている接続助詞 (あるいは助動詞) を誤り選択肢とした場合について，その適切さを評価する．

### 5.2 同義表現選択問題

同義表現選択問題とは，下記の問題例のように，問題文中の下線部と同義となる表現を選択肢から選択する問題である．

問題例 **2**

一級の試験に合格するために，
2000 字の漢字を 知っていなければならない．
1. 知っていなくてだめだ　2. 知っているわけではない

正用例文から同義表現選択問題を作成する際には，正解選択肢 (この場合は選択肢 1) については，元々の正用例文の動動詞 “なければならない” (意味分類: 禁止) と同義の助動詞 (例えば，この例の場合では，意味分類の細分類が “なければならない”

表 3: 機能表現学習用練習問題の選択肢自動生成における評価 (%)

| | 選択肢候補の判定結果 | 空欄補充問題 | 同義表現選択問題 | |
|---|---|---|---|---|
| | | 誤り選択肢 | 誤り選択肢 | 正解選択肢 |
| 接続助詞 | ○ | 74 (84/113) | 45 (71/158) | 91 (42/46) |
| | △ | 26 (29/113) | 23 (36/158) | 0 (0/46) |
| | × | 0 (0/113) | 32 (51/158) | 9 (4/46) |
| 助動詞 | ○ | 79 (117/149) | 65 (103/159) | 88 (22/25) |
| | △ | 21 (32/149) | 16 (26/159) | 0 (0/25) |
| | × | 0 (0/149) | 19 (30/159) | 12 (3/25) |

と同じ助動詞 “なくてはだめだ”) を用い，誤り選択肢については，接続助詞 “ために”(意味分類: 順接条件–因果関係) との呼応が可能な助動詞 “わけではない”(意味分類: 否定形による当為等・当為等の否定・不必要) を用いている．ここで，本節の評価では，正解選択肢の自動生成においては，接続助詞 (あるいは助動詞) の同義表現クラスを作成するための予備評価として，置換対象の接続助詞 (あるいは助動詞) と意味分類の細分類が同一の接続助詞 (あるいは助動詞) を正解選択肢とした場合について，その適切さを評価する．また，誤り選択肢の自動生成においては，接続助詞と助動詞の呼応を含む正用例文に対して，4. で作成した接続助詞・助動詞呼応可能性規則を参照し，正用例文中の助動詞 (あるいは接続助詞) と呼応可能とされている接続助詞 (あるいは助動詞) のうち，置換対象の接続助詞 (あるいは助動詞) とは意味分類の細分類が異なるものを誤り選択肢とした場合について， その適切さを評価する．

### 5.3 評価手順および結果

具体的には，以下の手順により，接続助詞と助動詞の呼応事例を正用例文として，接続助詞もしくは助動詞の練習問題の選択肢を自動生成する過程の評価を行った．

( 1 ) 3. で収集した接続助詞・助動詞呼応事例に対して，文長が 50 文字以下で，呼応関係にある接続助詞と助動詞の意味分類の大分類の組合せ (全 160 組) が分散される，という基準のもとで正用例文の候補を無作為に抽出した(注5)．さらに，抽出された呼応事例が練習問題の正用例文として適切かどうかを人手で判定して，評価用正用例文 20 文を選択した．

( 2 ) 空欄補充問題における誤り選択肢の評価:

評価用正用例文中の接続助詞 (あるいは助動詞) に対して，接続助詞・助動詞呼応可能性規則を参照し，正用例文中の接続助詞 (あるいは助動詞) と呼応不可能とされている助動詞 (あるいは接続助詞) を無作為に抽出して誤り選択肢とした場合について，以下の三通りのいずれであるかを判定する(注6)．

(○) 正用例文中の助動詞 (あるいは接続助詞) とは呼応不可能．
(△) 正用例文中の助動詞 (あるいは接続助詞) と呼応可能か否かの判定が困難．
(×) 正用例文中の助動詞 (あるいは接続助詞) とは呼応可能．

この判定結果を表 3「空欄補充問題 — 誤り選択肢」の欄に示す．この場合，(△) に判定されたものが一定数存在するものの，確実に呼応可能と判定されたものは存在せず，4. で作成した接続助詞・助動詞呼応可能性規則における判定結果がある程度信頼できると言える．また，適切な誤り選択肢 (○) を生成できる割合は，接続助詞で 74%，助動詞で 79%となる．

( 3 ) 同義表現選択問題における誤り選択肢の評価:

接続助詞と助動詞の呼応を含む正用例文に対して，接続助詞・助動詞呼応可能性規則を参照し，正用例文中の接続助詞 (あるいは助動詞) と呼応可能とされている助動詞 (あるいは接続助詞) のうち，置換対象の接続助詞 (あるいは助動詞) とは意味分類の細分類が異なるものを誤り選択肢とした場合について，以下の三通りのいずれであるかを判定する．

(○) 正用例文中の助動詞 (あるいは接続助詞) とは呼応可能．
(△) 正用例文中の助動詞 (あるいは接続助詞) と呼応可能か否かの判定が困難．
(×) 正用例文中の助動詞 (あるいは接続助詞) とは呼応不可能．

この判定結果を表 3「同義表現選択問題 — 誤り選択肢」の欄に示す．この場合，適切な誤り選択肢 (○) を生成できる割合は，接続助詞で 45%，助動詞で 65%となる (ただし，ここでの評価性能は，接続助詞・助動詞呼応可能性規則を見直すことによって，改善する余地があると思われる)．また，この場合，4. で作成した接続助詞・助動詞呼応可能性規則において，呼応可能な場合があり得ると判定された接続助詞・助動詞の組合せにおいても，問題文中の自立語との関係によって，呼応不可能となる場合が一定数存在することが分かる．このように，与えられた問題文の文脈下においては呼応不可能と判定された接続助詞・助動詞の組合せについては，逆に，上述の空欄補充問題における誤り選択肢として利用することができる．特に，4. で作成した接続助詞・助動詞呼応可能性規則において，呼応不可能と判定された接続助詞・助動詞の組は四分の一程度と少なく，自動処理において誤り選択肢とできる接続助詞・助動詞の組は限られるのが実情である．したがって，空欄補充問題における誤り選択肢のパターンを増やすためには，問題文中の自立語に応じて，誤り選択肢とできる接続助詞・助動詞の組の判定を半自動的に行う必要がある．

( 4 ) 同義表現選択問題における正解選択肢の評価:

接続助詞と助動詞の呼応を含む正用例文に対して，正用例文中の置換対象の接続助詞 (あるいは助動詞) と意味分類の細分類が同一の接続助詞 (あるいは助動詞) を正解選択肢とした場合について，以下の三通りのいずれであるかを判定する．

(○) 置換の前後で問題文の意味が不変である．
(△) 置換の前後で問題文の意味が不変であるか否かの判定が困難．
(×) 置換の前後で問題文の意味が変化する．

この判定結果を表 3「同義表現選択問題 — 正解選択肢」の欄に示す．この場合，(△) に判定されたものはなく，適切な正解

(注5): 語彙や表現の難しさの点を考慮すると，新聞記事中の文がそのまま日本語学習者用の練習問題に利用できるとは限らない．本論文で新聞記事中の文を評価の題材に用いているのは，あくまで練習問題の選択肢を生成する際の適切さを評価するためであり，そのまま練習問題として利用可能であるかどうかを直接評価しているわけではない．

(注6): 実際には，機械的に選択肢候補の接続助詞 (あるいは助動詞) との置換を行った場合には，この三通りの他に，「正用例文中の前後の語との接続において不適切な接続助詞 (あるいは助動詞) である」場合があり得る．ここでの評価では，そのような場合でも，当該接続助詞 (あるいは助動詞) について，大規模コーパスから例文を収集し，前後の語との間で形態素情報レベルの接続条件を記述しておくことで，前後の形態素を適切な活用形等に自動的に変形できると仮定している．以下の (3)，(4) の場合も同様である．

選択肢を生成できる割合は，接続助詞で 91%，助動詞で 88%となった．この割合は，上述の，誤り選択肢が適切な割合と比較してもやや高い値を示しており，意味分類の細分類が同一の接続助詞 (あるいは助動詞) の間では，高い率で表現の同義性が保たれていることが分かる．

## 6. 接続助詞・助動詞呼応の誤用自動訂正

ここでは，4. で作成した接続助詞・助動詞呼応可能性規則を用いて，接続助詞・助動詞呼応の誤用を自動訂正する方法の考え方について簡単に述べる．まず，接続助詞・助動詞呼応の誤用の検出においては，呼応可能性規則において呼応不可能とされている接続助詞と助動詞の組合せの呼応を用いた例文を誤用と判定する．次に，そのような誤用を訂正して，適切な表現の候補を提示するためには，作文者がどのような意味を表現するために呼応不可能な接続助詞と助動詞の組合せの呼応の表現を用いるに至ったかを推定し，作文者の意図を反映するような訂正表現を提示する必要がある．ここでは，そのような誤用訂正を行うための第一次近似的な解決法として，呼応可能性規則において呼応不可能と判定された接続助詞・助動詞の組合せについて，その組合せから推定される文の意味に最も近い意味を表現し，かつ，接続助詞・助動詞の呼応可能性の点からも適切な言い換え方を事前に用意しておくという方法をとる．

例えば，以下の例文 1 は，接続助詞 “とたん”(意味分類: 時間的関係–同時性) と助動詞 “てほしい”(意味分類: 要求・依頼) が呼応不可能なため，誤用を含む例文となっている．これは，接続助詞 “とたん” に関して「$S_1$ たとたん，$S_2$」という形式においては，前件 $S_1$ が既に完了済という意味合いを持つという制約が守られる必要があるが，例文 1 ではその制約が破られているためであると考えられる．これに対して，接続助詞 (的表現)“たらすぐに” や “と同時に” にはこのような制約がないので，例文 2 や例文 3 は誤用とはならない．

∗1. 試験の日程が発表されたとたん，連絡してほしい．
2. 試験の日程が発表されたらすぐに，連絡してほしい．
3. 試験の日程が発表されると同時に，連絡してほしい．

したがって，この例の場合には，接続助詞 “とたん” と助動詞 “てほしい” の間の呼応可能性に関する誤用に対して，接続助詞 “とたん” を言い換えることにより，元の文の意味に近い意味を表現し，かつ，誤用を解消する方法を事前に用意しておくことにより，接続助詞と助動詞の誤用の自動訂正が可能となる．

## 7. 接続助詞・助動詞単独の練習問題生成

5.1 および 5.2 では，接続助詞・助動詞の呼応可能性規則を利用して，接続助詞と助動詞の呼応事例を正用例文とする練習問題の作成法を述べたが，それらの方法は，接続助詞と助動詞の呼応事例を正用例文とする場合にのみ適用可能であり，接続助詞あるいは助動詞の任意の正用例文に対して適用できるわけではない．そこで，本節では，5.1 および 5.2 の考え方を発展させ，接続助詞あるいは助動詞の任意の正用例文に対して適用できる練習問題作成法の考え方について述べる．ここでの基本的な考え方は，二つの接続助詞 (あるいは助動詞) について，4. で作成した接続助詞・助動詞呼応可能性規則における特性が似ていれば，任意の例文における言い換え可能性も似ているという仮説に基づいている．以下に，空欄補充問題の例を挙げる．

**問題例 3**

ライバルも好調な (　　)，簡単に勝たせてくれるとは思えない．
1. だけに　　2. ところ

5.1 で述べたように，接続助詞 “だけに” と助動詞 “わけにはいかない” は呼応可能である一方，接続助詞 “ところ” と助動詞 “わけにはいかない” は呼応不可能である．したがって，接続助詞 “だけに” と接続助詞 “ところ” は，助動詞との呼応可能性において異なった特性を持っていると言える．このことに基づいて，下記の例文の空欄部分の言い換え可能性において，接続助詞 “だけに” と接続助詞 “ところ” は異なった特性を示すという仮説を立てることができる．したがって，以下の空欄補充問題において，元々の正用である接続助詞 “だけに” を正解選択肢とし，接続助詞 “だけに” とは異なった特性を示す接続助詞 “ところ” を誤り選択肢とすることができる．

同様に，5.2 で述べたように，助動詞 “なければならない”(意味分類: 禁止) と助動詞 “なくてはだめだ” は同義の関係にあるので，以下の同義表現選択問題においても，元々の正用である助動詞 “なければならない” に対して，“なくてはだめだ” を正解選択肢とすることができる．また，接続助詞との呼応可能性においては，助動詞 “なければならない” と “わけではない”(意味分類: 否定形による当為等・当為等の否定・不必要) は，いずれも，接続助詞 “ために”(意味分類: 順接条件–因果関係) とは呼応可能である．したがって，下記の例文の下線部分の言い換え可能性において，助動詞 “なければならない” と “わけではない” は同じ特性を示す仮説を立てることができ，実際，以下の同義表現選択問題において，言い換えはできるが “なければならない” とは異なった意味分類となる “わけではない” を誤り選択肢とすることができる．

**問題例 4**

一級の試験では，2000 字の漢字を 知っていなければならない．
1. 知っていなくてだめだ　2. 知っているわけではない

## 8. おわりに

本論文では，複文における接続助詞と助動詞の呼応可能性について調査を行った．二万組以上の接続助詞と助動詞の組合せのうち，約四分の三が呼応可能であることが分かった．さらに，接続助詞あるいは助動詞の練習問題の誤り選択肢を自動生成する過程において，接続助詞・助動詞の呼応可能性規則の評価を行い，5∼8 割程度の精度で適切な誤り選択肢が自動生成できた．

**謝辞:** 本研究は，一部，文部科学省 科学研究費 特定領域研究「高等教育改革に資するマルチメディアの高度利用に関する研究」(2000∼2002 年度) の援助を受けている．

### 文　献

[1] 黒川和也, 宇津呂武仁. コーパスおよび機能表現の分類体系を利用した日本語機能表現学習教材作成支援. 言語処理学会第 8 回年次大会論文集, pp. 431–434. 言語処理学会, 2002.
[2] 森田良行, 松木正恵. 日本語表現文型, NAFL 選書, 第 5 巻. アルク, 1989.
[3] 日本国際教育協会. 日本語能力試験出題基準. 凡人社, 1994.