日本語機能表現呼応可能性規則の作成および機能表現練習問題作成支援における評価* †

黒川 和也
豊橋技術科学大学 工学部 情報工学系
kazuya@cl.ics.tut.ac.jp

宇津呂 武仁
京都大学大学院 情報学研究科 知能情報学専攻
utsuro@i.kyoto-u.ac.jp

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．我々は，これまで，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 [黒川 02a]．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 [日本国際教育協会 94] 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した．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．

具体的には，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 [黒川 02b]．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．また，接続助詞あるいは助動詞単独の練習問題の誤り選択肢を自動生成する過程においても，4∼8 割程度の精度で適切な誤り選択肢が自動生成できた．

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 [黒川 02a] において我々が用いているものである．[黒川 02a] においては，[森田 89] の複合辞の階層的分類を参考にして機能表現を分類している．[森田 89] 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．この分類の概要を表 1 に示す．

3 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

*Judging Well-formedness of Collocation of Japanese Functional Expressions and its Evaluation in Exercises Authoring

†本研究は，一部，文部科学省 科学研究費 特定領域研究 (A)「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」研究課題「音声言語処理技術と学習者モデルを用いた語学学習システムの研究」(2000∼2002 年度，研究代表者: 豊橋技術科学大学 工学部 中川 聖一 教授) の援助を受けている．

表 1: 機能表現の分類 (機能表現例)

助詞の働きをする表現 (273 種類)				
格助詞タイプ	係助詞タイプ	副助詞タイプ	接続助詞タイプ	終助詞タイプ
資格・立場 (をもって，でもって), 対象・関連 (について ,にかんして)，全 9 分類，63 種類	定義 (とは，というのは)，主題化 (といえば ,といって) ,全 2 分類，30 種類	強調 (という，といった)，限定・非限定 (に限って ,にかぎり) ,全 6 分類，28 種類	同時性 (や否や，が早いが)，仮定 (としても，にしても)，全 10 分類，129 種類	伝聞 (とのこと，ということ)，願望・勧誘 (かな)，全 9 分類，46 種類
助動詞の働きをする表現 (177 種類)				
義務・当然 (なければならない)，アスペクト (ずにいる，てしまう) 全 16 分類				

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．その結果，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った [黒川 02b]．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た．

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

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．
2. その他の場合，机上実験により判定．

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 [黒川 02b]．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．

5 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 [黒川 02a] 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 節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，元々の正用例文の助動詞 “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” 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

表 2: 機能表現学習用練習問題の誤り選択肢自動生成における評価 (%)

	選択肢候補の判定結果	接続助詞・助動詞呼応練習問題		接続助詞・助動詞単独の練習問題	
		空欄補充	同義表現選択	空欄補充	同義表現選択
接続助詞	○	76 (667/874)	53 (534/1000)	70 (278/400)	59 (237/400)
	△	24 (207/874)	10 (99/1000)	1 (5/400)	3 (11/400)
	×	0 (0/874)	37 (367/1000)	30 (117/400)	38 (152/400)
助動詞	○	85 (597/698)	63 (626/1000)	43 (260/600)	78 (470/600)
	△	15 (101/698)	15 (152/1000)	2 (13/600)	18 (113/600)
	×	0 (0/698)	22 (222/1000)	55 (327/600)	3 (17/600)

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 節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

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．3 節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．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 100 文を選択した．

5.3.1 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．

- (○) 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 2「接続助詞・助動詞呼応練習問題: 空欄補充問題」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 節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 76%，助動詞で 85%となる．

5.3.2 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 2「接続助詞・助動詞呼応練習問題: 同義表現選択問題」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 53%，助動詞で 63%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 節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 節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．

6 接続助詞・助動詞単独の練習問題生成

本節では，5.1 節および 5.2 節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 節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 節で述べたように，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 節で述べたように，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

以上の考え方に基づいて，接続助詞あるいは助動詞単独の正用例文に対して，練習問題の選択肢生成過程の評価を行った．空欄補充問題においては，接続助詞・助動詞呼応可能性規則における特性が最も似ていない 10 個の機能表現を，また，同義表現選択問題においては，接続助詞・助動詞呼応可能性規則における特性が最も似ている 10 個の機能表現を，それぞれ誤り選択肢の候補とした．評価用正用例文としては，接続助詞を含む正用例文 40 文，助動詞を含む正用例文 60 文の合計 100 文を用いた．この評価結果を表 2 「接続助詞・助動詞単独の練習問題」の欄に示す．この結果から分かるように，呼応可能性の特性が最も似ていない表現を選ぶか，逆に，呼応可能性の特性が最も似ている表現を選ぶかの間で，接続助詞の場合は 70%(空欄補充での○) 対 38%(同義表現選択での×) という差となり，助動詞の場合は 43%(空欄補充での○) 対 3%(同義表現選択での×) という差となった[1] ．したがって，上述の仮説が一定の割合で成立していると言える．

7 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．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，接続助詞・助動詞の呼応可能性規則の評価を行った．

参考文献

[黒川 02a] 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).

[黒川 02b] 黒川和也, 宇津呂武仁：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, 信学技報, TL2002-33∼41, pp. 19–24 (2002).

[森田 89] 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).

[日本国際教育協会 94] 日本国際教育協会：日本語能力試験出題基準, 凡人社 (1994).

[1] 助動詞の場合は，無作為に抽出した評価用例文に単文が一定数含まれており，必然的に，誤用の選択肢が生成しにくく，逆に，言い換え可能な選択肢の生成が容易になってしまった．今後，評価用例文を全て複文とするなどの操作を施した後に，再評価をする必要がある．