# コーパスを利用した日本語機能表現学習教材作成支援* †

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
{kazuya,utsuro}@cl.ics.tut.ac.jp

## 1　はじめに

近年，国際化の進展によって多くの日本人が海外で生活し，また逆に多くの外国人が日本で生活するようになった．このため，非母語言語の習得支援体制の確立が急務となっており，近年では， 計算機による語学学習のためのマルチメディア語学教材の研究や開発なども盛んになってきた．その一方で，近年，各種の言語資源の電子化が進みつつあり，また，それらを利用する自然言語処理技術も向上してきた．さらに，計算機の性能も格段に向上し，大量の言語資源を高速に検索することが可能となっている．したがって，これまで完全に人手に頼っていた語学学習用教材作成の作業を，計算機を用いて効果的に支援することが可能となりつつある．完全に人手に頼った教材作成と比較して，計算機に用いて教材作成を支援するアプローチの利点として，個々の学習者のレベルや誤りパターンにあわせて，教師が最適な練習問題を作成できる可能性が広がるという点が挙げられる．さらに，今後，教材作成の過程の自動化が実現できれば，自習の際にも，個々の学習者のレベルや誤りパターンにあわせて動的に練習問題を作成することも可能となり，学習者の学習効率を大きく上げることが期待できる．

このような背景のもとで，本研究では，計算機上の自然言語コーパス (正用コーパス) を利用して，日本語学習者のための教材開発を支援する方法について研究を行なう．特に，本研究では，日本語の文法習得において重要な役割を持つ機能表現に焦点をあて，機能表現の練習問題を，日本語コーパスから半自動的に作成する方法について研究を行なう．また，学習者の母語と日本語が対訳になったコーパスから，学習者の母語による翻訳情報が一部付与された日本語機能表現練習問題を作成する事例も示す．

具体的には，本研究では，日本語能力試験 [日本国際教育協会 94] で出題の対象となる接続・修飾・文末などの機能表現を対象とする．さらに，日本語コーパス，あるいは，日本語と英語が対訳になったコーパスから，各機能表現の用例を検索するための語彙・品詞パターンを記述し，この検索パターンを用いて各機能表現の用例の検索を行なう．また，新規の練習問題の作成においては，正用コーパスから検索された機能表現の用例に対して，正用の機能表現とは意味分類の異なる機能表現への置き換えを行なった誤用例を生成し，これらの正用と誤用を組み合わせて新規の練習問題とする．さらに，作成された練習問題では，学習者の理解度を向上するために，各選択肢の機能表現を含む (対訳) 用例を検索しておき，必要に応じてこれらを提示することも可能である．

## 2　日本語機能表現練習問題

### 2.1　日本語能力試験

日本語能力試験 [日本国際教育協会 94] は，日本国内および国外において，日本語を母語としない者を対象として，日本語能力を測定・認定することを目的として，財団法人日本国際教育協会と国際交流基金が共催で実施している試験である．科目の類別は，「文字・語彙」，「聴解」，「読解・文法」の三類別からなり，一級から四級の四つのレベルで構成されている．

### 2.2　機能表現練習問題

日本語能力試験の問題のうち，「文法」の問題は，基本的な文型・活用事項，助詞・指示語・疑問詞，接続・修飾・文末などの機能表現，敬語，などの問題から構成される．このうち，本研究では，一・二級の文法問題の主要な問題である接続・修飾・文末などの機能表現の練習問題作成支援を対象とする．本稿では，特に，一級の文法問題集 [田中 92] に挙げられている機能表現リスト (接続，修飾，その他表現として 25 分類 382 種類，文末表現として 27 分類 287 種類 — 抜粋を表 1 に示す) を対象とする．

また，[田中 92] では，文法問題の出題形式として，主として以下の三つの形式を挙げている．

---

*Corpus-based Authoring of Exercises for Learning Japanese Functional Expressions

†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」の援助を受けている．本研究に関して，多くの貴重なコメントを頂いた，豊橋技術科学大学 情報工学系 中川聖一教授をはじめとする同特定領域研究のメンバー，ならびに，(株) リコー ソフトウェア研究所 成田真澄氏に感謝する．

表 1: 機能表現の分類 (機能表現例) 一覧

| 接続，修飾，その他表現 (25 分類) |
|---|
| 題材・対象・関連 (ということは)，例示・説明・同格 (といった)，資格・立場・状態・視点 (からいうと)，<br>推移・相関 (につれて)，起点・終点・範囲 (を通じて)，対応・割合・基準 (に対して)，対比・代替・仮想対比 (と比べて)，<br>挙例・列挙・枚挙 (といい ～ といい)，時・場所・状況 (て以来)，仲介・手段・根拠・基準 (をたよりに)，<br>強調・程度 (とばかりに)，限定・除外・不都合 (を除いて)，付加・非限定 (と同時に)，推測・推量 (ようで)，<br>可能・不可能 (がたく)，理由・因果関係・根拠 (する以上)，意志・決意 (つもりで)，目的 (ために)，<br>伝聞・引用 (とのことで)，婉曲 (みたいで)，結果・継起 (～ たうえで)，仮定条件 (くらいなら)，確定条件 (てはじめて)，<br>逆接仮定条件 (としても)，逆接確定条件 (にもかかわらず) |
| 文末表現 (27 分類) |
| 理由・因果関係 (おかけだ)，義務・当然・必然・必要・勧告 (なければならない)，<br>禁止・制止・当然の否定・不必要 (てはいけない)，可能・不可能 (ことができる)，許容・許可・自責 (てもよい)，<br>意志・意図・決定 (つもりだ)，意向強調 (ないではいられない)，自然成立 (ことになる)，自発・強制 (ざるをえない)，<br>断定・主張・説明 (にちがいない)，伝聞・引用 (そうだ)，結論・説明・帰結 (というわけだ)，傾向 (がちだ)，<br>推量・推測・推定・予測 (かもしれない)，比況 (みたいだ)，婉曲 (のではないか)，希望・願望 (たいものだ)，<br>提案・勧告・勧誘 (てほしい)，疑問提示・反語 (ではないか)，命令・要求・依頼・期待 (てくれ)，限定 (よりほかない)，<br>程度 (に過ぎない)，経験・回想・習慣 (ことがある)，行為の授受 (てくれる)，アスペクト (たところだ)，<br>反実アスペクト (するところだった)，仮定・比況設定 ((仮に) としよう) |

問題例 **1** 私のプライバシーに関することですので，
詳しいことは (　　)．

1. 申し上げられます
2. 申し上げかねます
3. 申し上げるわけではありません
4. 申し上げないわけにはいきません

問題例 **2** 先生，すみません．あした父の友人が日本に来るんです．その人は日本語が全然分からないので 迎えに行かなければいけないんです．

1. 迎えに行かないわけではないんです
2. 迎えに行くわけではないんです
3. 迎えに行くわけにはいかないんです
4. 迎えに行かないわけにはいかないんです

問題例 **3** 大雪の ため あしたの試験は延期します．

1. 母の ため 明るい色のスカーフを買いました．
2. あなたは勉強の ため と言って，
そのテレビを買ったんでしょう．
3. 病気の ため 試験が受けられませんでした．

問題例 1 は，空欄に最も適した選択肢を選ばせる問題，問題例 2 は，下線部と同じ意味の選択肢を選ばせる問題，問題例 3 は，複数の用法を持つ機能表現について，問題文の下線部と同じ用法の選択肢を選ばせる問題である．このうち，本稿では，同義性の厳密な判定の取り扱いを必要とせず，比較的作成が容易な問題例 1 の形式の練習問題の作成支援を対象とする．

# 3　機能表現練習問題作成支援

コーパスを利用して機能表現練習問題作成の過程を支援する枠組の概要を以下に示す．

1. 学習の対象とする機能表現を選定する[1]．
2. 機能表現の用例検索用コーパスを選択する．
3. 用例コーパスから，学習対象の機能表現の用例を検索する．
4. 検索された用例のうちの一つを問題文とする．
5. 学習対象の機能表現に対して，誤り選択肢となる機能表現をいくつか決定し，誤り選択肢を生成する．
6. 問題文，正解の選択肢，誤り選択肢を組合わせて，新規の練習問題とする．

これらの各過程の詳細，および練習問題作成の事例について，以下の各節で順に説明する．

## 3.1　用例検索用コーパス

日本語能力試験の出題基準 [日本国際教育協会 94] においては，一・二級の機能表現の用例調査の対象として，新聞・文学作品・雑誌・新書の類などの用例が用いられている．近年，これらのテキストの電子化が進みつつあることから，一・二級レベルの機能表現の用例を検索用コーパスとして，それらの電子化テキストが利用できると考えられる．ただし，各級で出題が認められている漢字・語彙の範囲についても詳細に定められており，例えば，一級では，語彙数が一万となっているので，練習問題に用いる用例についても，これらの範囲を逸脱しないという条件を満たす必要がある．

また，3.3 節に示すように，機能表現検索用コーパスとして，学習者の母語と日本語が対訳になったコーパスを利用することにより，学習者の母語による翻訳情報が一部付与された日本語機能表現練習問題 (図 1

[1] 本稿では，学習の対象とする一つの機能表現に対して，誤り選択肢の機能表現をいくつか生成するという設定を仮定している．ただし，用法の似ている二つ (またはそれ以上) の機能表現の違いを習得するという設定においても，本稿の枠組の構成要素の多くはそのまま利用できる．

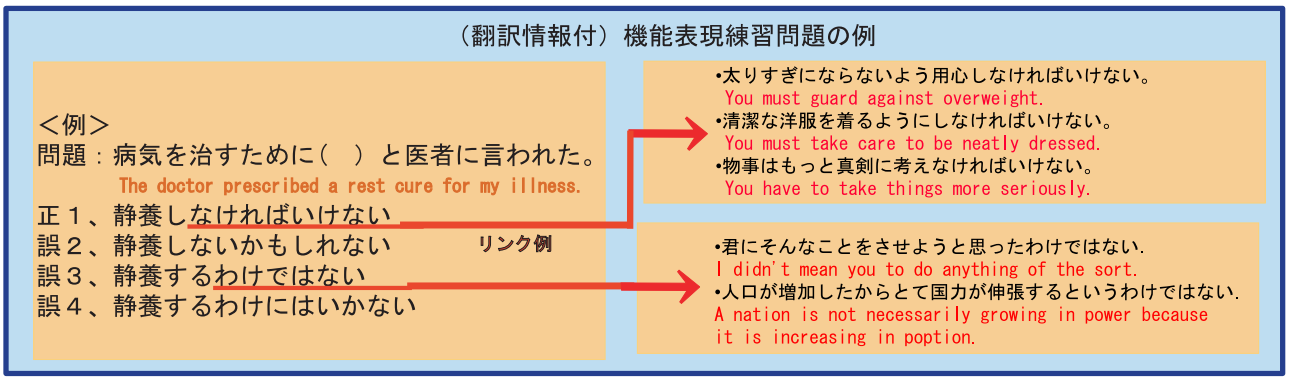


図 1: (翻訳情報付) 日本語機能表現練習問題の例

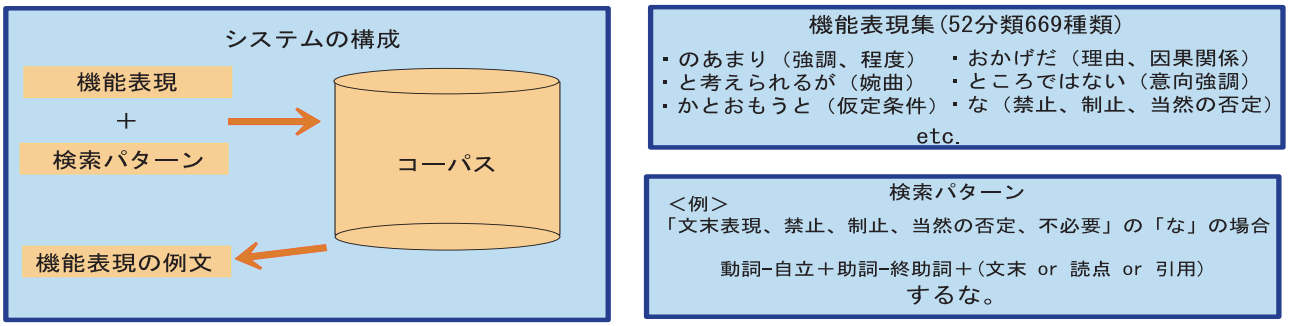


図 2: 日本語機能表現の検索

参照 — 各選択肢の機能表現から対訳用例へのリンク付) を作成することができる．ただし，この場合，学習者の母語と日本語が対訳になったコーパスとしてどのようなものが利用可能であるかが問題となる．現時点で，日本語との対訳コーパスとして，ある程度まとまった数の機能表現を網羅できる規模のものが存在するのは，日本と英語が対訳になったコーパスで，しかもその多くは，英和・和英辞典などを電子化したオンライン辞書中の対訳例である．そこで，本稿では，英和・和英などのオンライン対訳辞書の対訳例文を利用して，英訳付日本語機能表現練習問題の作成を支援する事例を示す[2] [3] ．

### 3.2 用例検索

機能表現の用例を検索する際には，用例文を形態素解析 [松本 99] した結果に対して，あらかじめ用意しておいた検索パターンを用いて各機能表現の用例を検索する．検索パターンの作成の際には，まず，機能表現と用例文の文字列照合により，用例候補文を検索する．次に，用例候補文を形態素解析した結果を，機能表現の文字列の周囲の語彙・品詞パターンでクラスタリングし，このクラスタリング結果を人手で選別することにより，機能表現の周囲の語彙・品詞により記述された検索パターンを得る．これまでのところ，文字数の少ない機能表現約二十種類について検索パターンを作成した．文字数の多い機能表現ほど，文字列照合によっても適切な検索結果が得られることが多いので，数百種類の機能表現について検索パターンを作成することは可能であると考えている．例として，文末表現のうち，「禁止・制止・当然の否定・不必要」に分類されている終助詞の「な」の検索パターンを図 2 に示す．

### 3.3 誤り選択肢の生成

学習対象の機能表現に対して，誤り選択肢となる機能表現をいくつか決定し，誤り選択肢を生成する過程について，図 3 の例を用いて説明する．図 3 では，「義務・当然・必然・必要・勧告」に分類されている文末表現「なければならない」に対して，誤り選択肢を生成する様子を示している．ここで，誤り選択肢となる機能表現を選定する基準として，本稿の方法では，一級の文法問題集 [田中 92] に挙げられている機能表現リストの

[2] 日本語能力試験の一級レベルからすると，対訳辞書の対訳例の日本語文はやや平易である．特に，新聞・文学作品・雑誌・新書の用例と比較すると，参照可能な文脈の長さに制限があるため，(一級レベルで必要とされる) 微妙な用法の区別の理解度を測定するような難易度の高い問題が作りにくいという問題がある．したがって，より適切な難易度で，かつ，学習者の母語による翻訳情報が付与された練習問題の作成を実現するためには，新聞・文学作品・雑誌・新書などの対訳テキストの収集が必要である．

[3] 日本語学習者の多くの割合を占めるのはアジア諸国出身者であり，英語以外のアジア諸国語 (たとえば，中国語，韓国語など) を母語とする者の割合が多いのが現実である．したがって，現実には，英語以外のアジア諸国語と日本語との対訳コーパスが必要とされる割合が多いことが予想される．

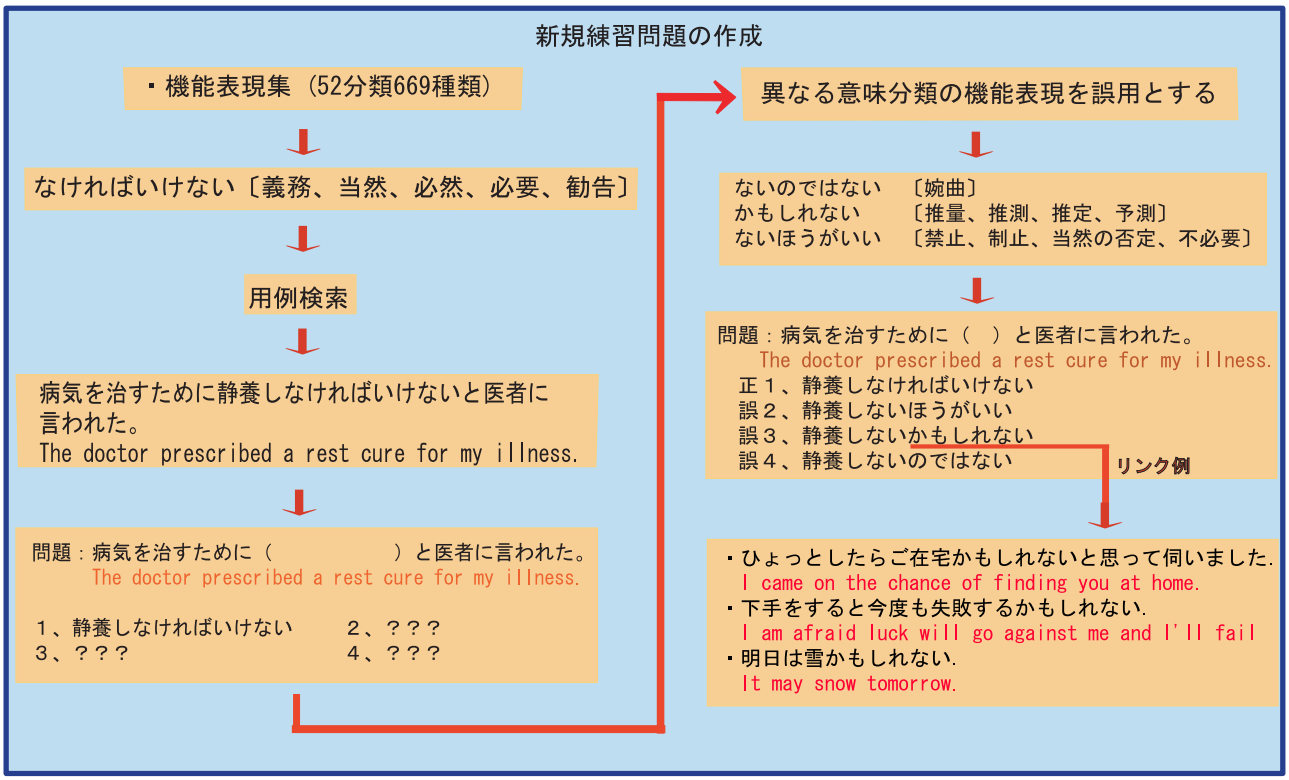


図 3: 新規練習問題の作成

52 分類を利用する．具体的には，学習対象の機能表現の意味分類 (この場合，「義務・当然・必然・必要・勧告」) とは異なる意味分類を持つ機能表現を，誤り選択肢の候補としてシステムが問題作成者に提示し，問題文の空欄に当てはめても不自然でなく，かつ，正解の機能表現とは意味の異なるものを問題作成者が選択する．その際に，正解文の自立語部分と誤り選択肢の機能表現の間の接続可能性は，誤り選択肢の機能表現の検索パターンを用いて検査することができる．

ここで，今回利用した一級文法問題集 [田中 92] の機能表現の 52 分類は，一次近似的な最も粗いものであり，より難易度の高い微妙な用法の区別を扱うためには，より詳細な分類 (例えば，[森田 89] の分類など) を採用する必要がある．また，そのような詳細な分類を用いて，機能表現の用法間の意味的な近さを定義できれば，選択肢の機能表現の間の意味的な近さを調節することにより，学習者のレベルに応じて問題の難易度を柔軟に調整することが可能となると考えられる．さらに，今後，誤り選択肢の生成過程を自動化するためには，日本語教育の専門家が作成した練習問題の誤り選択肢の規則性を分析したり，日本語学習者の実際の誤用例 [寺村 90, 市川 97, 大曽 99] の誤りパターンを分析するなどして，誤り選択肢の生成規則について検討する必要がある．

## 4 おわりに

本稿では，コーパスを利用して，日本語の機能表現の練習問題作成過程を支援する方法について述べた．現時点では，まだ，小規模な実験を行なっている段階であるので，今後，用例検索用コーパスの適切性，学習対象の機能表現の網羅性，日本語教育の専門家が作成した練習問題との比較，実際の学習者を対象とした評価，翻訳情報の有効性，など様々な観点から評価を行なっていく予定である．

## 参考文献

[市川 97] 市川保子：日本語誤用例文小辞典, 凡人社 (1997).

[松本 99] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 浅原正幸：日本語形態素解析システム『茶筌』version 2.0 使用説明書 第二版, NAIST-IS-TR99012 (1999).

[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).

[大曽 99] 大曽恵美子：日本語学習者の作文コーパス: 電子化による共有資源化, 文部省科学研究費 基盤研究 (A)(1) 研究成果報告書 (1999).

[田中 92] 田中望 (監修), 友松悦子, 宮本淳, 和栗雅子：日本語テスト問題集 — 文法編 —, 凡人社 (1992).

[寺村 90] 寺村秀夫：外国人学習者の日本語誤用例集, 文部省科学研究費 特別推進研究「日本語の普遍性と個別性に関する理論的及び実証的研究」, 分担研究「外国人学習者の日本語誤用例の収集，整理及び分析」資料 (1990).