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* †

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，各種の言語資源の電子化が進みつつあり，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．計算機を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．本研究では，計算機上の自然言語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の研究を行い，とくに，日本語の文法習得において重要な役割を持つ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の研究を行う．具体的には，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 [黒川 01]．本研究では，計算機を用いた日本語文法学習支援の過程を，

1. 日本語機能表現の分類体系の構築，分類体系を用いた練習問題作成支援の研究
2. 学習者の誤り分析，および，分類体系を用いた練習問題作成支援技術の運用の統合による，実用的練習問題作成支援システム・機能表現学習用 CALL システムの研究

に分けてとらえ，特に 1 に焦点をあてる．

2　日本語機能表現の分類体系の構築

機能表現練習問題を作成する際は，機能表現の意味分類だけでなく，意味は異なるが統語的接続形式がほぼ同等の機能表現の組が必要となる．そこで，本論文では，意味および統語的接続パターンの両方を考慮して，日本語機能表現を表 1 のように分類する．

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，[森田 89] 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，[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 分類・273 種類) と，助動詞の働きをするもの (13 分類・150 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．

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，機能表現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．

- 非文末型体言接続
  - 例：　格助詞タイプ・・・私 として
  - 係助詞タイプ・・・私 といえども
  - 副助詞タイプ・・・私 という
- 文末型体言接続
  - 例：　終助詞タイプ・・・私 かな
  - 助動詞タイプ・・・私 とは限らない
- 非文末型用言接続
  - 例：　格助詞タイプ・・・する にあたって
  - 係助詞タイプ・・・する となると
  - 副助詞タイプ・・・する ことは
  - 接続助詞タイプ・・・する や否や
- 文末型用言接続
  - 例：　終助詞タイプ・・・する ものだ
  - 助動詞タイプ・・・する ことはいけない

3　機能表現の組の分類

次に，機能表現の階層的分類体系をもちいて練習問題を作成するために，機能表現の組を以下の 4 タイプに

*Authoring Exercises of Japanese Functional Expressions utilizing Corpora and Syntactic/Semantic Classification of 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」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: 意味分類および統語的接続パターン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.)，全 9 分類，63 種類
	用言接続	時・場所・状況，同格，全 2 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 分類，30 種類
	用言接続	主題化，全 1 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 分類，25 種類
	用言接続	強調，全 1 分類，6 種類
接続助詞タイプ	用言接続	同時性 (や否や，が早いが，そばから，etc.)，継起 (あげく，ところ，etc.)，相関 (に従い，につれて，etc.)，順接-仮定 (かぎりは，ようものなら，etc.)，順接-確定 (てみると，てみたら，etc.)，因果関係 (からには，おかげで，etc.), 逆接-仮定 (としても，にしても，にしたって，etc.)，逆接-確定 (からといって，にもかかわらず，etc.)，対比 (のに対して，と同時に，etc.)，反復・並立・添加 (まま，ばかりか，etc.)，全 10 分類，96 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 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 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 分類，24 種類
	用言接続	禁止 (てはいけない，etc.)，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.)，当為等の否定・不必要 (ものではない，とは限らない，etc.)，可能・不可能 (ことができる，わけにはいかない，etc.)，許容・許可 (てもいい，てもかまわない，etc.)，意志・決定 (ようとする，つもりだ，etc.)，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.)，推量・推測・推定 (かもしれない，はずだ，etc.)，適当・願望・提案・勧誘・勧告 (ばいい，てほしい，etc.)，要求・依頼 (てくれ，てください，etc.)，限定 (よりほかない，までだ，etc.)，程度 (に過ぎない，だけのことはある，etc.)，経験・回想・習慣 (ことがある，ものだ，etc.)，伝聞 (とのことだ，etc.)，行為の授受 (てやる，てもらう，etc.)，アスペクト (ずにいる，てしまう，ずにしまう，etc.)，全 13 分類，150 種類
合計		助詞タイプ: 36 分類・273 種類，助動詞タイプ: 13 分類・150 種類

分類した．

- タイプ 1： 置き替えによって文の意味が不変の組

  例 : 病気を治すために静養し (なければならない/しなくてはだめだ)．

  この例の場合は，“なければならない” を，他の機能表現 “しなくてはだめだ” などに置き換えても文の意味が変わらない．

- タイプ 2： 置き替えによって文の意味が変わるが日本語として許容できる組

  例 : 病気を治すために静養 (しなければならない/するわけではない)．

  この例の場合は，“なければならない” を，他の機能表現 “わけではない” と置き換えた場合，日本語として許容できるが，文の意味が変化する．

- タイプ 3： 置き替えによって日本語として許容できなくなる組

  例: 先生すみません．明日，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ならない/*迎えに行くに違いない) んです．

  この例の場合は，“なければならない” という機能表現を，“に違いない” に置き換えると，日本語として許容できなくなってしまう．

- タイプ 4： タイプ 1∼3 以外で，言い替えによって，練習問題の素材にすることが適当でない組．このタイプについては，作業過程でタイプ 1∼3 のいずれにも適さない例があれば認定する．

4 機能表現の組の分類を利用した練習問題作成

前節の機能表現の組の分類を用いれば，タイプ 1 とタイプ 2 の違いを利用した同義表現選択問題と，正用と

タイプ 3 の違いを利用した空欄補充問題の二種類の形式の練習問題が作成できる．[黒川 01] では，空欄補充問題のみ扱ったが，問題文の文脈によって，誤用の選択肢でも文として意味が通じてしまうという問題点があった．そのために，問題作成の対象となる機能表現が限定されていたが，同義表現選択問題を用いれば，問題作成の対象となる機能表現の範囲が広がる．以下にそれぞれについて述べる．

4.1 タイプ 1 とタイプ 2 の違いを利用した練習問題の形式

問題例 **1** 病気を治すために 静養しなければならない．

1. 静養しなくてはだめだ
2. 静養するわけではない

上記の問題例 1 は，問題文中の下線部 “静養しなければならない” と同義となるものを選択肢から選択する問題である．この問題では，問題文の “なければならない” とタイプ 1 の関係がある “なくてはだめだ” を正解選択肢として，タイプ 2 の関係にある “わけではない” を誤り選択肢として用いている．

4.2 正用とタイプ 3 の違いを利用した練習問題の形式

問題例 **2** 先生すみません．明日，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（　　）んです．

1. 迎えに行かなければならない
2. 迎えに行くに違いない

上記の問題例 2 は，空欄に適当な表現を選択肢から選択する問題である．正解選択肢については，元々の文の機能表現 “なければならない” をそのまま用いて，誤りの選択肢は，問題文の機能表現とタイプ 3 の関係にある機能表現 “に違いない” を用いている．

5 機能表現練習問題作成の流れ

図 1 に，機能表現練習問題作成の流れを示す．まず，練習問題の対象とする機能表現を検索キーとして，正用の例文を正用データベースから検索する．次に，正用の例文に対して，正用中の機能表現との間でタイプ 1∼ タイプ 4 となる機能表現の組を認定する．この認定は，練習問題作成者が行う．機能表現の組の認定においては，機能表現全体を六つのグループ

- 非文末型 { 体言/用言 } 接続の二グループ
- 文末型 { 体言/用言 } 接続
  ×{ 終助詞タイプ/助動詞タイプ }
  の四グループ

(1)

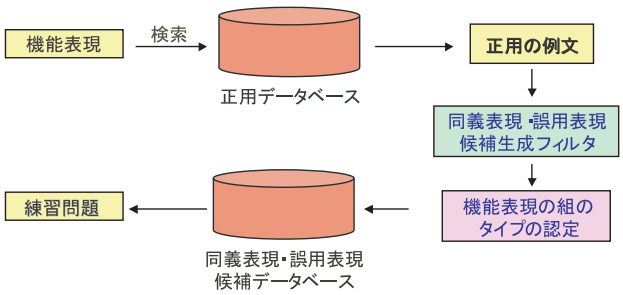

図 1: 機能表現練習問題作成の流れ

に分類し，各々のグループ内の機能表現の組に限定して認定を行う．ここでは，機能表現の組認定の作業効率を上げるために，機能表現の組の候補に対して適当なフィルタを設定することもできる．認定された機能表現の組は正用例文とともに「同義表現・誤用表現データベース」に蓄積される．実際に練習問題を作成する際には，機能表現の組認定とは別のフェーズを設け，「同義表現・誤用表現データベース」中の正用例文および同義表現・誤用表現の候補を参照して，自動もしくは半自動で機能表現練習問題が生成される．

6 機能表現練習問題作成過程の評価

本論文の機能表現練習問題作成過程の有効性を評価するために，表 2 に示す全 40 種類の機能表現について，各々一文ずつ正用の例文を検索し，各例文の正用機能表現に対して，(1) 式に示す六グループのうちの該当グループの範囲内の全機能表現との間で，タイプ 1∼ タイプ 4 の組の認定を行った．表 2 には，機能表現の分類ごとに，認定された機能表現組の数の内訳 (タイプ 1∼ タイプ 4) を示す．全体では，4905 組の機能表現組のタイプの認定を行ったが，本研究で作成した作業用インタフェースを用いた場合の作業効率は，おおよそ 500 組/時間程度であった．

図 2 に，タイプ 1∼ タイプ 4 の機能表現組の割合の分布のうち，特徴的なものの抜粋を示す．まず，全機能表現組における分布においては，「同義表現・誤用表現候補生成フィルタ」として，正用の機能表現と同一意味分類の機能表現のみを候補とするフィルタを用いた場合の分布 (a)，および，用いない場合 (b) を示す．両者の比較より，タイプ 1 の同義表現の割合は，同一意味分類の範囲の限定した場合の方が圧倒的に大きいことが分かる．このことから，タイプ 1 の同義表現を効率よく認定することが目的の場合には，同一意味分類の機能表現のみを候補として機能表現組の認定を行えばよいことが分かる．次に，非文末型のうち，体言接続 (c)，および，用言接続 (d) の機能表現に限定した場合の分布を示す．(b) の全体での平均と比較すると，

表 2: 評価用機能表現一覧・評価用正用例文数・タイプ 1∼ タイプ 4 の機能表現組数

非文末型		
格助詞タイプ	体言接続	資格・立場 (からすると)，対象・関連 (に関して)，仕手・仲介・手段・根拠・原因 (によると，を通じて)，対応 (によっては)， 全 5 文，タイプ 1: 47 組，タイプ 2: 171 組，タイプ 3: 462 組，タイプ 4: 55 組
	用言接続	時・場所・状況 (に際して)， 全 1 文，タイプ 1: 3 組，タイプ 2: 49 組，タイプ 3: 23 組，タイプ 4: 90 組
係助詞タイプ	体言接続	主題化 (ときたら)， 全 1 文，タイプ 1: 10 組，タイプ 2: 31 組，タイプ 3: 106 組，タイプ 4: 0 組
	用言接続	全 0 文
副助詞タイプ	体言接続	強調 (のあまり)，限定・非限定 (を問わず)，添加 (のみならず)， 全 3 文，タイプ 1: 11 組，タイプ 2: 5 組，タイプ 3: 415 組，タイプ 4: 10 組
	用言接続	全 0 文
接続助詞タイプ	用言接続	同時性 (や否や)，継起 (あげく)，相関 (につれて)，順接-仮定 (とすると)，順接-確定 (としたら)，因果関係 (からには，ものだから), 逆接-仮定 (にしても)，逆接-確定 (からといって)，全 9 文， タイプ 1: 68 組，タイプ 2: 160 組，タイプ 3: 183 組，タイプ 4: 732 組
文末型		
終助詞タイプ	体言接続	疑問・問いかけ・確認 (ではないか，だって)，断定・強調 (というものだ)， 全 3 文，タイプ 1: 5 組，タイプ 2: 4 組，タイプ 3: 3 組，タイプ 4: 28 組
	用言接続	感動・詠嘆・驚異 (ものだ)，疑問・問いかけ・確認 (かい，かな，たっけ)，反語・反駁・否定・非難・後悔 (ものを，のくせに) 願望・勧誘 (ないか)，回想 (たっけ)， 全 8 文，タイプ 1: 30 組，タイプ 2: 87 組，タイプ 3: 117 組，タイプ 4: 190 組
助動詞タイプ	体言接続	全 0 文
	用言接続	禁止 (てはだめだ)，義務・当然・当為・必然・必要・勧告・主張 (なければならない，よりほかない)，当為等の否定・不必要 (わけがない，はずがない)，許容・許可 (てもいい)，意志・決定 (ようとする)，適当・願望・提案・勧誘・勧告 (ばいい)，要求・依頼 (てくれないか，てほしい)， 全 10 文，タイプ 1: 85 組，タイプ 2: 751 組，タイプ 3: 736 組，タイプ 4: 238 組
合計		全 40 文，タイプ 1: 259 組，タイプ 2: 1258 組，タイプ 3: 2045 組，タイプ 4: 1343 組

体言接続ではタイプ 3 の組が多く，用言接続ではタイプ 4 の組が多くなっているが，いずれのタイプの組も，正用の例文中の機能表現を誤用の機能表現に言い換えることにより，非文となる場合に相当する．非文末型の機能表現の場合，正用の例文が複文に近い構文の場合が多く，文中の他の自立語・機能語との共起関係の制約が強いために，言い換えにより非文となり易い．接続助詞タイプのような用言接続の場合には，特にその傾向が強く，練習問題としての適格性をも満たさずタイプ 4 となることが多い．さらに，文末型用言接続の助動詞タイプの場合について，正用の例文が単文の場合 (e)，および，複文の場合 (f) の分布を示す．単文の場合は，文中の他の自立語・機能語との共起関係の制約が弱く，言い換えても非文とならないため，タイプ 2 の組の割合が多い．一方，複文の場合は，文中の他の自立語・機能語との共起関係の制約が弱く，言い換えにより非文となり易いため，タイプ 3 の組の割合が多い．

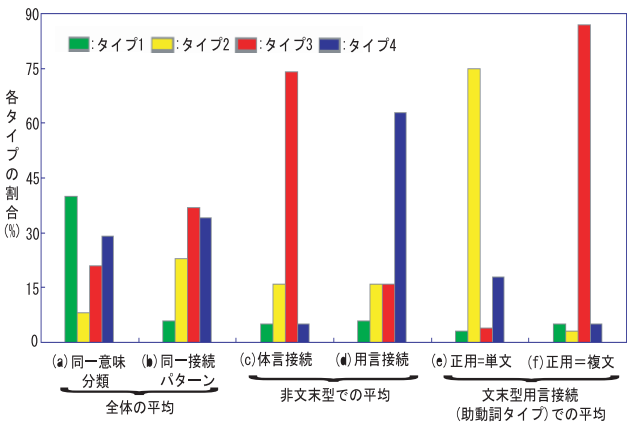

図 2: タイプ 1∼ タイプ 4 の機能表現組の割合の分布 (統語的接続パターンごと，抜粋)

7 おわりに

本稿では，日本語機能表現練習問題作成支援に必要な日本語機能表現の分類体系，および，それを用いた機能表現練習問題作成の方法について述べた．また，日本語機能表現練習問題作成過程の評価を行い，機能表現の統語的接続パターンごとに，各タイプの機能表現組の割合の分布についてその傾向を分析した．

参考文献

[黒川 01] 黒川和也, 宇津呂武仁 : コーパス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 7 回年次大会論文集, pp. 367–370, 言語処理学会 (2001).

[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).