

# 論文から取り出した基本表現を用いた作文練習システム

西村 涼 谷口 友哉 渡辺 靖彦 岡田 至弘

龍谷大学 理工学部 情報メディア学科

{r\_nishimura,t\_taniguchi}@afc.ryukoku.ac.jp, {watanabe,okada}@rins.ryukoku.ac.jp

## 1 はじめに

論文、レポート、説明書などを正しく書くためには、多くの文章を読み、その分野の用語と表現を正しく理解し、知識として得た用語と表現をくりかえし使用することが重要である。しかし、時間的・空間的な制約をはじめ、さまざまな理由で効率よく学習するのはむずかしい。例えば、大学における用語と表現の学習機会は、

- 用語と表現との接触は専門講義科目
- 用語と表現の使用は実験演習科目

とわかれていて、作文技術の学習をむずかしくしている。このため、学習者への時間的・空間的制約をゆるめ、それぞれの専門分野の用語と表現を効率よく学習する方法の1つとしてe-Learning システムへの期待が高まっている。

e-Learning システムでは、

**Distance Learning** ネットワークを介して講義を行う遠隔学習

**Web Based Training(WBT)** WWW技術をベースに、ブラウザ上に表示される教材や講義を見ながら行う学習

などが代表的な例としてあげられる[e 白書 03]。このほかにも、どのようなタイミングでシステムが用いられるかによって、

**同期型** 教室での授業のように、教師が学習者を常に手助けするシステム

**非同期型** 教師がいなくても学習ができるシステムに分類することができる。

本研究では、龍谷大学における初年度生向けの計算機基礎演習科目を対象に作成した、基本表現の学習を目的とした作文練習システムについて述べる。作成したシステムであつかう基本表現は、レポートや論文などを作成するとき多用される4種類(1. 比較、2. 対比、3. 類似、4. 原因・理由)である。この基本表現を論文から取り出す方法についても述べる。

## 2 論文からの基本表現の抽出

本研究では、以下の4種類の情報を表現する文を論文から取り出し、作文練習システムで基本表現を学習するための例文として利用する。

**比較** 同質の2つ以上のものを量的に比べる

**対比** 2つの異なるものを質的に比べる

**類似** 2つのものに共通点があることを示す

**原因・理由** 事物の変化をひきおこすものや事物の成り立っているみちすじ・わけを示す

例文を抽出する対象として、言語処理学会第12回年次大会の予稿集に含まれる論文を用いた。基本表現を含む例文は、以下の手順で取り出す。

### step 1 [形態素解析および係り受け解析]

予稿集に含まれる論文317件を対象に、形態素解析および係り受け解析(並列構造解析)を行う。形態素解析にはJUMAN [1]、係り受け解析(並列構造解析)にはKNP [2]を用いた。

### step 2 [比較、類似、原因・結果の基本表現を含む例文の抽出]

比較、類似、原因・結果の基本表現を含む文は、step 1の形態素解析結果から取り出す。図1に本実験で取り出す比較、類似、原因・結果の基本表現を示す[山崎 02]。

### step 3 [対比の基本表現を含む例文の抽出]

対比の基本表現を含む文は、図1に示すように、助詞「が」の前後にある「～は～」の並列構造を手がかりにしてstep 1の係り受け解析(並列構造)解析から取り出す。図2は、助詞「が」の前後の「～は～」が並列構造である場合で、図1の対比の基本表現を含む文に対する規則で取り出せる。

規則	基本表現	意味	例
比較 1	～より～のほうが～	同質の2つのものを比較	鉄より鉛のほうが比重が大きい
比較 2	～は ～と比較すると ～と比べると ～と比較して ～と比べて ～よりも ～より	比較する対象の一方に重みがある	氷は水と比較すると密度が小さい
対比	～は～が、～は～ (← 並列構造 →)	2つの異なるものを質的に比較	ナトリウムは陽イオンになるが、塩素は陰イオンになる
類似 1	～と～は ～も～も	両方について説明	空気と水はともに流体である
類似 2	～は ～とともに ～と同様に ～と同じく	類似例で説明	この物質は、金属と同様に酸で劣化する
類似 3	～は ～と等しい ～に等しい	質量も同じ	1気圧は水銀柱 76cm と等しい
類似 4	～は～と同等である	同じに扱える	荷重に対するゴムの変形挙動はばねと同等である
類似 5	～は～と等価である	質のみが同じ	この回路は(式1)と等価である
原因・理由 1	～ ため、 (こと)により、 ので、 結果、	原因・理由を述べ、次に結果を述べる	剛性が不足したため、固有振動数が低下した
原因・理由 2	～。 このため、 これにより、 この結果、	原因・理由を述べ、次の結果を述べる	剛性が不足した。このため、固有振動数が低下した
原因・理由 3	～は～ためである。	結果を述べ、次に原因・理由を述べる	温度が上昇したのは、反応が進んだためである。
原因・理由 4	～。これは～	結果を述べ、次に原因・理由を述べる	温度が上昇した。これは反応が進んだためである。

図 1: 抽出する基本表現

表 1 に、基本表現を含む文の抽出結果を示す。表 1 に示すように、比較の基本表現ではない文を誤って取り出している失敗が多い。これは、以下の例のような文を比較の基本表現として取り出したからである。

(例) 図 2 より「find the way」の場合、文章化した場合 のほうが 文章化しない場合より 2 倍近く評価度が上がった。

「より」と「のほうが」の前にあらわれる語が同質のものかどうかを判定しなければならない。しかし、シソーラスを用いても、以下の例のように、「より」と「のほうが」の前にあらわれる語が同質であると判定するのがむずかしい場合も多かった。

(例) つまり、どのように談話の展開を行っても、談話全体のレベルでは依頼文の丁寧度が低い場合

より 高い丁寧度の依頼文を有する談話のほうが 丁寧に感じられている。

対比の基本表現ではない文を誤って取り出した失敗も多い。しかしその中には、対比の情報を表現しているものの、基本表現の例文としては適切でないと判断して誤りと判断した例も多い。例えば以下の例は対比の情報を表現しているが、「が」の前には「には」があり、後には「では」が用いられているので、抽出結果は誤りと判定した。前の「には」が「では」であれば、この例は基本表現の例文を正しく取り出したと判定した。

(例) 最初の複合名詞 には 2 種類の係り受け構造しかないが、次の複合名詞 では 5 種類の係り受け構造がある。

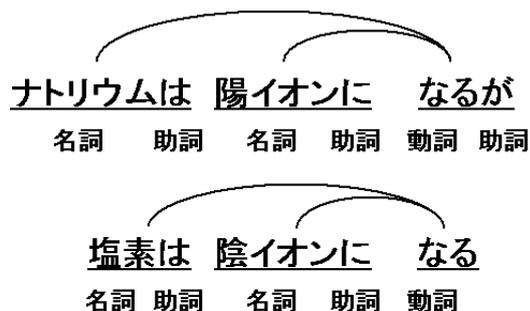


図 2: 助詞「が」の前後にある「～は～」の並列構造の例

表 1: 基本表現を含む文の抽出結果

規則	正しく抽出	誤って抽出	合計
比較 1	5	4	9
比較 2	90	870	960
対比	102	97	199
類似 1	5	21	26
類似 2	52	36	88
類似 3	2	0	2
類似 4	0	0	0
類似 5	2	0	2
原因・理由 1	1960	189	2149
原因・理由 2	80	0	80

### 3 基本表現学習システム

作成したシステムの概要を図 4 に示す。作成したシステムは、

1. 学習者 (ユーザ) は、学習したい基本表現を選択する。
2. ユーザが学習したい基本表現を図 3 の (a) に示すような問題形式にして示す。
3. ユーザは図 3 の (b) のような形式で問題に答えると、システムが正解かどうかを自動的に判定する。

システムを構成するモジュールの機能と内容やデータベースの内容について以下に示す。インタフェースには Web ブラウザを用い、システムは Perl で実装されている。データベースには、MySQL を用いた。学習したい基本表現の問題は、データベースに簡単に追加できるようになっている。

基本表現選択モジュール 学習者 (ユーザ) が指定し

(a) 問題  
 全損失関数 、観測された2つ組のラベル変数にのみ重みが増える 、点損失関数  2つ組のラベル変数のうち、どちらかが観測されていれば重みが増えるようにパラメータが更新される。

(b) 解答  
 全損失関数  では、観測された2つ組のラベル変数にのみ重みが増える  が、点損失関数  では 2つ組のラベル変数のうち、どちらかが観測されていれば重みが増えるようにパラメータが更新される。

図 3: 対比の問題と解答の例

た基本表現を受け取り、問題生成モジュールに送る。

問題生成モジュール ユーザが指定した基本表現を学習できるような問題をデータベースから取り出した基本表現を含む例文から 3 つ作成する。生成した問題は、問題提示モジュールに送る。

問題提示モジュール 問題生成モジュールが生成した問題をユーザに提示する。ユーザは提示された問題を解答する。ユーザの解答は、正解判定モジュールに送る。

正解判定モジュール データベースに格納されている基本表現を含む例文とユーザの解答を比較して、正解かどうかを判定する。判定結果を正解提示モジュールに送る。

正解提示モジュール ユーザの解答が正解かどうかを正解判定モジュールの判定結果にしたがって、ユーザに提示する。

比較/対比/類似/原因・理由 2 章で述べた手法を用いて抽出した基本表現を含む例文のうち、正しく抽出できた例文が格納されている。

### 4 システムの利用例

作成したシステムで「対比」の基本表現の問題を学習する場合の画面を図 5 に示す。データベースには、2 章で述べた手法で抽出した 199 例の「対比」の例文のうち正しく抽出できた 102 例の例文が格納

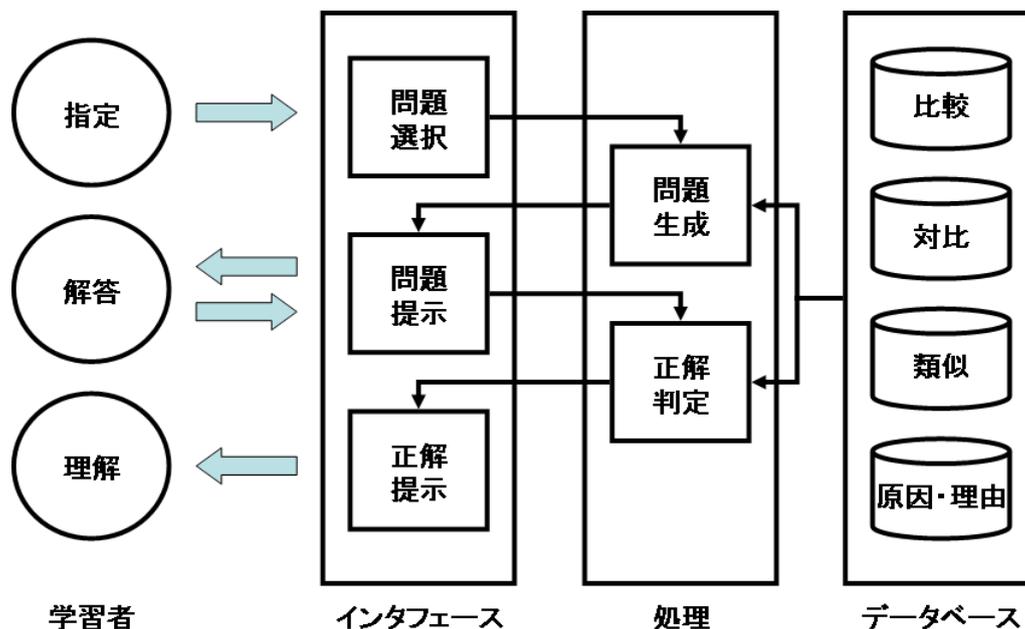


図 4: システムの概要

## 大学生のための日本語練習サイト

問題はランダムに3つ出題されるので、テキストボックスに適切な答えを入力してください。

### 比較 (comparison)

- つまり、どのように談話の展開を行っても、談話全体のレベルでは依頼文の丁寧度が低い場合、 高い丁寧度の依頼文を有する談話  丁寧に感じられている。
- 代名詞  非代名詞  頻出するので話者も受話者も慣れ、認知的コストが低下する。
- この表からわかるように、三角形モデルとU字型モデル  他と  良い性能を示している。

[問題を变える](#) [基本表現選択画面に戻る](#) [解答する](#)

図 5: 問題を提示している画面の例

されている。このデータベースからランダムに3つ例文を選びだし、基本表現の部分を穴埋め問題として学習者に提示している。

学習者は、図5のような穴埋め問題を解くことで、科学技術文書やレポートなどを正しく書くために必要な表現を効率的に身につけることができる。

- [1] 黒橋, 河原: 日本語形態素解析システム JUMAN version 5.1 使用説明書, 京都大学, (2005)
- [2] 黒橋, 河原: “日本語構文解析システム KNP version 2.0 使用説明書.”, 京都大学, (2005).

## 参考文献

- [山崎 02] 山崎, 富田, 平林, 羽田野: 科学技術日本語案内, 慶応義塾大学出版会, (2002).
- [e 白書 03] 先進学習基盤協議会編著: eラーニング白書 (2003/2004 年度版), オーム社, (2003).