

# AI チュータの実現に向け：誤用例文コーパス データの構築と誤用文修正知識の習得

相川孝子(マサチューセッツ工科大学)

高橋哲朗 (Fujitsu Laboratories of America, 現)

本研究では、人工知能を用いた日本語学習支援ツール「AI チュータ」の実現に向け、その第一ステップとして行なった誤用文<->修正文ペア・コーパスデータの構築過程、またそのコーパスを基に習得した文法規則の分析を行なう。この AI チュータは日本語学習者（特に初級～中級レベル）が作文を書く際によく犯す文法的な誤りを自動的に発見し、それらの修正方を提示することを主な機能とする。本稿では、まず誤用修正のための知識習得を目的に行なったペア・コーパスデータの構築過程、そしてそのデータを基に習得された文法規則などを紹介する。次に、修正文の中に見られた日本語教師間での修正方の揺れについて言及し、どのような言語パターンが教師間での修正方の揺れを引き起こすのか考察する。後半は、ペア・コーパスデータから習得された規則の問題点に触れながら、今後どのようなアプローチで AI チュータのシステムの向上に努めていったらいいか考えていく。

キーワード：人工知能、クラウドソーシング、誤用修正、自然言語処理、

## 1 はじめに

テクノロジーの活用が各領域で進む中、語学教育に於いても様々なオンライン上での個人学習環境が提供されている。ただ、多くのオンラインツールやアプリは、まだ学習者とシステムとの双方向性（“interactivity”）に欠け

る。そこで、語学学習に不可欠な双方向性とは何か、特に学習者の能動的学習（“active learning”）の支援に注目しながら、日本語学習支援ツール「AI チュータ」の開発に取り組んだ。AI チュータは、日本語学習者が文章を書く際に犯す文法的誤りをリアルタイムで修正する機能を持つ支援ツールであるが、開発の際、最大の課題になったのが、学習者の誤りを自動的に判定し、それを修正する日本語の知識をいかに構築するかということであった。この課題に対し、まず考えられる方法は、手書きによるルール作成であるが、この方法では、学習者が犯す誤用の予期できないため、誤用判定ができない。そこで、日本語教師に誤用例文の作成とその修正という2つの作業を委託し、誤用文<->修正文ペア・コーパスを構築、そのコーパスデータから修正知識を習得するアプローチを取った。本稿では、この委託作業の枠組みを「Teacher Sourcing」と呼ぶ。

## 2 誤用文<->修正文ペア・コーパスの構築:Teacher-Sourcing

### 2.1 誤用文<->修正文ペア・コーパスの作成

日本語学習者の誤用例文のコーパスに関しては、過去、複数の研究がなされているが(寺村 1990, 大曾(他)1997, Mizumoto, et. al 2011, 大山(他)2012, 望月 2012, 李 2012, 大山(他)2013)、本研究とこれらの先行研究との違いは、本研究では誤用文修正のための知識収集に、Teacher Sourcing の枠組みを使い、誤用文<->修正文ペア・コーパスデータを作成したことにある(高橋・相川 2016)。また、ただ単に言語データとしてのコーパスを作るのではなく、構築されたコーパスデータから、誤用修正知識を自動的に習得し、その知識を AI チュータのシステムに再利用させるという点でも、従来の研究とは異なる。

この Teacher Sourcing は、過去2回に渡り行なったが、作業委託を依頼したのは、現在北米で日本語を教えている教師である<sup>1</sup>。表1は、Teacher Sourcing の期間及びその参加者のプロフィールを示す。

表 1 : Teacher Sourcing 参加者のプロフィール

	期間	K-12/Higher Ed.	Native Languages
初回	10/1-11/30, 2015	高校教師 (2) 大学教師 (8)	日本語母国者 (8); 英語母国者 (1); 日英バイリンガル(1)
2回目	7/1-8/31, 2016	高校教師 (4) 大学教師 (6)	日本語母国者(6); 英語母国者(4)

Teacher Sourcing の過程としては、まず参加者の教師から初級～中級レベルの学生がよく犯す文法的誤用を含む例文 200 文の作成を依頼した。次に、図 1 にあるような Web インタフェースを用意し、Teacher Sourcing によって収集された誤用文を Web 上で修正する作業を委託した。Web インタフェース上での作業過程は、まず、表示された誤用文の中で誤っている個所の範囲を指定 <1>。次に誤用文編集のための画面がポップアップ表示され、そこに修正文を入力し保存 <2>。最後にその修正パターンを用いて修正された結果が元の画面に表示される <3>というものである。



図 1 : 誤用例文修正のワークフロー

上の作業過程に従い、初回の Teacher Sourcing では、各教師に 400 文の誤用例文の修正を依頼し、それぞれの誤用文が少なくとも 2 人の教師によって修正されるようにした。また、2 回目の場合は、各教師に 800 文の修正を依頼し、それぞれの誤用例文が少なくとも 4 人の教師によって修正されるようにした。

## 2.2 修正パターン規則の習得:初回 Teacher Sourcing の結果

2 回に渡る Teacher Sourcing で集積された誤用文<->修正文ペア・コーパスデータを利用し、誤用文の修正知識習得を試みたが、基本的には、誤用文と修正文の表層文字列の差を修正パターン規則とし、それを AI チュータの

システムにフィードする仕組みである。(1)は、表層文字列を基にした修正規則の具体例を、図2は、修正規則の習得過程を図式化したものである。

- (1) \*飲みて->飲んで  
\*行くじゃない->行かない  
\*行きましたです->行きました

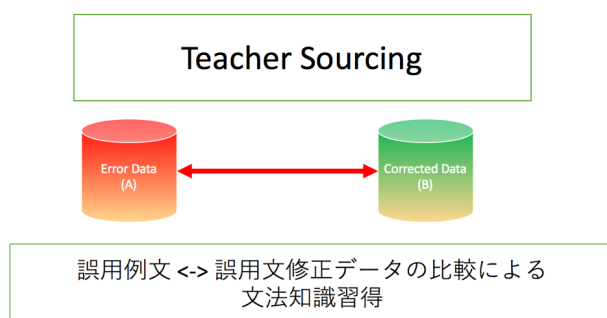


図2：修正パターン規則の習得過程

初回の Teacher Sourcing では、集積された誤用文<->修正文ペアデータ(4,000文ほど)の中で、修正パタンの異なり数が3,766件あり、このうち事例をまたいで再利用可能なもの(即ち、2回以上使われた規則)は、73件ほどあった。これは、データサイズを大きくすれば、更に多くの再利用可能な修正規則が習得できることを示唆した。同時に、初回 Teacher Sourcing の経験から、形態素解析を施すことにより、修正規則の一般化の必要性も感じた。例えば、(2)に見られるような修正パターン例は、(3)のような抽象化が可能である。

- (2) \*元気かった->元気だった  
\*きれいかった->きれいだった  
\*便利かった->便利だった
- (3) [名詞-形容動詞語幹]かった->[名詞-形容動詞語幹]だった

また、初回のデータから、修正パターンは、教師によって異なるという現象も発見した。この教師間による「修正パタンの揺れ」については、第3章

で詳しく触れるが、実際 Teacher Sourcing から得られたデータを見ること  
によって、一体どんな文法項目が教師間での修正の揺れを生じさせるのか、  
また日本語を母国語とする教師と英語を母国語とする教師の間で修正パタン  
の違いがあるかなど、新しい課題も発見することができた。2 回目の  
Teacher Sourcing は、こうした新しい課題も念頭におき、行なった。

### 2.3 第 2 回 Teacher Sourcing の結果とその考察

2 回目の Teacher Sourcing によって集積された誤用文は、初回同様 2,000  
文ほどであったが、2 回目の場合、教師間での修正パタンの揺れについて  
細かく分析したかったため、各教師に 800 文の修正作業を委託した。これ  
により、収集された誤用文<->修正文のペアデータのサイズが増加された訳  
だが、本来なら 8,000 件(10 人 x 800 文)の修正規則があるはずだったが、  
重複、修正もれなどがあつたため、延べ数 7,992 件になった。この 7,992  
件の修正規則のうち、異なり数は 5,298 件。このうち、1 度だけ使われた  
規則の数は 3,768 件であった。つまり、1,530 個の規則が 2 回以上使われ、  
4224 件(7,992 件-3768 件)の誤用例文の修正を行なっているわけである。  
これは修正パタンがロングテールであることを示唆しており、出現頻度の  
高い一定数の修正規則が作成できれば、それらの規則により比較的多くの  
割合の誤用文を自動的に修正できるという可能性がある。そこで、2 回目  
に集積された 7,992 件の修正規則がランダムな順序で作成されたと仮定し、  
規則の再利用によって新たな規則の追加が不要になる様子をシミュレーシ  
ョンした。図 3 は、その結果を示す。

実線は比較対象のために規則の再利用をせずに毎回修正した場合の作成規  
則の数を表わす( $y=x$ )。これに対し破線は、すでに得られた規則の再利用を  
考慮し、新規に追加される規則の数のみをプロットした曲線である。図 3  
に於いて、修正規則の追加に伴い、次第に実線と破線の幅が広がっている  
ことから、追加すべき規則の数が鈍化していることが分かる。この結果か  
ら、教師が新しい種類の誤りを修正するたびに新しい修正規則が作られ、  
既知の誤りに対しては AI チュータが教師の代わりに修正できるようになる

可能性を示すことができた。但し、誤り修正のバリエーションは非常に大きいため、すぐにはこの曲線が水平にはならないことが予想される。

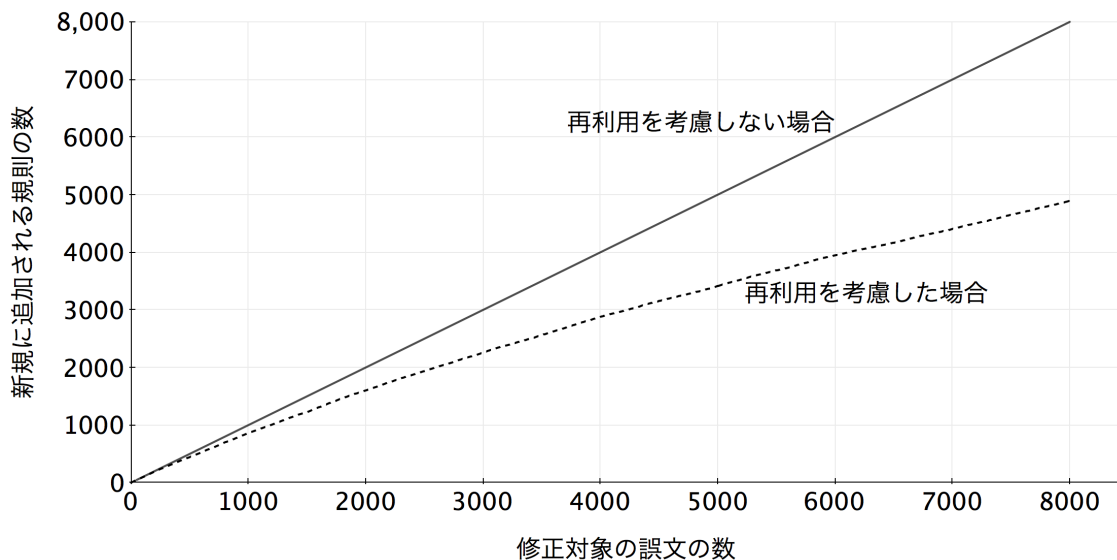


図 3：修正規則のロングテール化

### 3 教師間での修正パタンの揺れ

本章では、Teacher Sourcing で収集されたデータから発見された教師間での「修正パタンの揺れ」を下の二つの観点から捉えてみた。

- 修正指定範囲の違い
- 修正パタンの違い

修正指定範囲というのは、教師が誤用文を修正する際、どこからどこまでを修正範囲として指定するかという問題である（図 1 にある<1>の過程を参照）。例えば、(4)の誤用文に対して、ある教師は、(4a)のように述部だけを修正範囲として指定したが、ある教師は、(4b)のように目的語も含めた部分を指定していた。

(4) \*私のしゅみは、映画を見るです。（下線部が修正指定範囲を示す）

(4)a. \*私のしゅみは、映画を見るです。

[修正文]->私のしゅみは、映画を見ることです。

(4)b. \*私のしゅみは、映画を見るです。

[修正文]->私のしゅみは、映画を見ることです。

これに対し、修正パタンの違いは、実際の修正文字の違いを指す。(4)の例の場合は、修正指定範囲は異なるが、実際の修正パターンには、違いがない。一方、(5)のような場合は、修正指定範囲も修正パターンも異なる例である。

(5) \*私は、五人家族があります。(下線部が修正指定範囲を示す)

(5)a. \*私は、五人家族があります。

[修正文]->私は、五人家族です。

(5)b. \*私は、五人家族があります。

[修正文]->私の家族は、五人います。

### 3.1 修正指定範囲の違い

まず、初回と2回目のデータにおける教師間の修正指定範囲の違いを統計的に調べてみた。表2は、2回のデータにおける修正指定範囲一致率を示す。

表2：修正指定範囲一致率

	完全一致	部分一致	不一致
初回 Teacher Sourcing	32%	65%	3%
2回目 Teacher Sourcing	33%	62%	5%

表2の数値からも分かるよう、教師同士による修正範囲の一致率は、大変低いことが分かる。ただ、ここで一つ注目したいのは、教師の修正指定範囲が一致した場合、同じ修正パターンをとる割合は75%という高い割合になっていた。即ち、教師間で、一旦修正箇所が一致すれば、同じような修正方法を取るという傾向である。

### 3.2 修正パタンの違い

教師間に於ける修正パタンの違いに関しては、次の2つの観点から考察してみた。

- 修正パタンの揺れと文法項目との関係
- 日本語母国語教師と英語母国語教師との修正方の違い

### 3.2.1. 修正パタンの揺れと文法項目との関係

まず、どんな文法項目が含まれた時、教師間での修正パタンの揺れが多く見られたかという問題を取り上げる。この問題に対しては、構築した誤用文<->修正文ペア・コーパスから、修正パタンの異なりが多く見られた文をとりあげ、次にそれらの文にどんな文法項目が含まれているかを調べてみた。本稿では、その中で、最も頻度数の高かったもの3項目とその例を表3に挙げる。

表3：修正方法の揺れを引き起こす文法項目

文法項目	誤用例文とその修正例
疑問詞	(6) *何プレゼントをあげますか。 [修正1] どんなプレゼントをあげますか。 [修正2] 何のプレゼントをあげますか。
動詞（テ形）	(7) *オーディオを聞いて、帰ります。 [修正1] オーディオを聞いてから、帰ります。 [修正2] オーディオを聞きながら、帰ります。
自動詞、他動詞	(8) *ドアが開けています。 [修正1] ドアを開けています。 [修正2] ドアが開けてあります。 [修正3] ドアが開いています。

では、どうして上記の文法項目が含まれると修正パタンの揺れが起こりやすいのか。その原因の一つとして、これらの文法項目を含んだ文は、筆者の意図が掴みにくいということが考えられる。例えば、(6)の疑問詞の誤用文の場合、筆者の意図が、(i) プレゼントのタイプを聞いているのか、または、(ii) プレゼントの目的を聞いているのか、この誤用文からだけでは分からない。もし、筆者の意図が(i)の場合は、修正1が、(ii)の場合は、修正2が考えられるが、コンテキストがないため、筆者の意図が分からない。故に、教師間での修正パタンの揺れが生じると仮定される。

(7)の場合も、この誤用文からだけでは、筆者の意図が分からない。動詞（テ形）の用法は、(i)時間的前後関係を示す継起用法もあれば、(ii)付帯状況的に後続の動作を修飾する付帯用法もある(吉永 2012)。もし、筆者の意図が(i)の場合は、修正1が適当であるが、(ii)の場合は、修正2の方が



適当となる。同様に、(8)のような自動詞・他動詞を含んだ誤用文も、一文からだけでは筆者の意図が読み取れないので、教師間での修正パタンの揺れが生じると仮定される。このように、初級～中級の一見単純そうな誤用文も筆者の意図が分からない場合は、教師の側もその修正方にバラツキをもってしまうことがよく分かる。

### 3.2.2 日本語母語話者教師と英語母語話者教師との違い

次に日本語を母国語とする教師（ネイティブ）と英語を母国語とする教師（ノンネイティブ）とで、修正方に何か違いがあるかを探ってみたが、一つの傾向として、ネイティブ同士、またはノンネイティブ同士の場合は、その修正パタンの一致率が高い。特に、ノンネイティブ同士の場合、非常に高い一致率を示す。表4は、修正パターン一致率を示したものである。

表4：修正パターン一致率の違い

(a) ネイティブ同士の一致率	36.97
(b) ノンネイティブの一致率	39.37
(c) ネイティブとノンネイティブの一致率	32.22

ノンネイティブ教師間での高い一致率は、彼らが日本語学習者としての文法知識（規範文法）を基に、即ち同じ文法規則を基に修正を行っているため、一致率が高くなると仮定できる。

また、ノンネイティブの教師が行なった修正文には、英語からの言語転移（language transfer）の影響がみられる場合がある。例えば、(9)に見られるような誤用文に対してネイティブの教師は、(9a)にあるような「象は鼻が長い」文型（三上 1960）、または(9b)のような所有格を使つての修正方をとるが、ノンネイティブの教師による修正には、(9c)のように英語からの影響が見られる<sup>2</sup>。

(9) [誤用文]\*シンさんは、長い髪がです。

(9)a. [修正1]シンさんは、髪が長いです。[ネイティブ]

(9)b. [修正 2] シンさんの髪は、長いです。[ネイティブ]

(9)c. [修正 3] ??シンさんは、長い髪があります。[ノンネイティブ]

また、(10)にあるような存在文のパタンを含んだ誤用文にもネイティブとノンネイティブの差がはっきり現れてくる。

(10) [誤用文]\*私は、家族が五人あります。

(10)a. [修正 1] 私は、五人家族です。[ネイティブ]

(10)b. [修正 2] ??私は、五人家族がいます。[ノンネイティブ]<sup>3</sup>

こうしたネイティブとノンネイティブの修正パタンの差は、文法性の容認度の問題(ナガラ 1988)にも繋がり、今後言語学的観点からもう少し深く分析していきたい。

## 4 規則の限界とその対策

本章では、現在の AI チュータが直面している修正規則の問題点、また、それらの問題点に対しての今後の対応策について述べる。

### 4.1. カバレッジの限界

まず、最初にとりあげられる問題は、カバレッジの問題である。これは、現在の修正規則が、コーパスデータに含まれる誤用文・修正文に基づく表層文字配列の変換ルールであるため、コーパス内にある誤用例のケースは対処できるが、データ内にないケースは対処できないという限界である。例えば、(11)のような形容動詞「きれい」を含む誤用例は、コーパスに入っているため、対処できるが、全く同じ誤用パターンである(12)のケースは、コーパス内に形容動詞「妥当」を含む誤用例がないため、対処できない。

(11) \*あの部屋はきれいなです。->あの部屋はきれいです。[対処可能]

(12) \*この問題は妥当なです。->\*この問題は妥当なです。[対処不可]

この問題に対しては、第2章でも述べたよう、形態素解析を施し、表層文字配列に頼っている修正規則をより一般化させることを考えている。例えば、(11)のような修正パタンのケースは、我々のコーパス内に多くある。(13)はその一部を示したものだが、こうした例から、形態素解析の情報を利用し、(14)にあるような一般的な修正規則を学ぶのは、決して不可能ではない。

- (13) \*元気なです->元気です; \*有名なです->有名です;  
\*ヘルシーなです->ヘルシーです; \*好きなです->好きです
- (14) [名詞-形容動詞語幹]なです->[名詞-形容動詞語幹]です

#### 4.2. コンテキストの配慮

次に、どんなコンテキストで、どの規則が適用可能であるか判断ができないという問題である。例えば、(15)の修正規則は、(15a)のコンテキストでは適用されるべきであるが、(15b)の場合、適用されるべきではない。同様に、(16)の規則も(16a)では適用されるべきであるが、(16b)の場合、適用されるべきではない。

- (15) 食べてあります->食べています
- (15)a. \*ごはんを食べてあります。->ごはんを食べています。
- (15)b. ごはんが食べてあります。->\*ごはんが食べています。
- (16) 買ってくれました->買ってあげました
- (16)a. \*山田さんはトムに本を買ってくれました。->山田さんはトムに本を買ってあげました。
- (16)b. 山田さんは私に本を買ってくれました。->\*山田さんは私に本を買ってあげました。

これは、現在のAIチュータが入力文のコンテキストをチェックできないため、修正規則として学ばれた文字配列が入力文にあった場合、一律にその規

則を適用してしまうからである。この問題は、(15b)や(16b)に見られるように、正文を非文にしてしまうという危険も齎し、真剣に取り組んでいかなければいけない。ただ、「コンテキストの理解」は、自然言語処理の分野でも難題であり、今後我々にとっても、大きなチャレンジとなると考えられる。この問題に対しては、現在、機械学習、特にニューラルネットワークの利用を考えている。というのは、そもそもこの誤用修正のタスクは、機械翻訳のタスクに似ているからである。機械翻訳の場合は、ある言語から他言語への変換作業であるが、誤用修正の場合は、誤った日本語から正しい日本語への変換作業と言ってよい。最近の機械学習、ニューラルネットワークなどは、従来、難関とされていたこのコンテキストの理解に対処できるとされ、機械翻訳、音声、また対話システムなどの分野に渡って大きな成果をあげている。その意味でも、こうした新しいテクニックが、コンテキストを考慮しながらの誤用修正に大きな効果を与えてくれることを期待する。

#### 4.3. ユーザーフィードバック

修正規則のカバレッジとコンテキストの問題は、自然言語処理、機械学習など「コンピューターの力」を借りながら、改善に努めると同時に、「人間の力」、即ち、ユーザーからのフィードバックを利用しながら、AI チュータの修正規則の洗練も試みたい。これは、ユーザーからのデータを得るというクラウドソーシングのプロセスであるが、先に紹介した Teacher Sourcing とは、方法が全く異なる。このユーザーフィードバックの利用のために、現在 AI チュータには、「サムズアップ・ダウンの機能」と「規則の編集機能」が備わっている。

サムズアップ・ダウンの機能は、ユーザーに AI チュータから提案された修正が正しいか正しくないかを判断してもらい、修正規則の洗練に繋げることを目的としている。上でも述べたように、現在のシステムから提案される修正候補には、正しいものあれば、間違っただけのものがあるが、ユーザーは、図4にあるような「サムズアップ・ダウン」の機能によって、提案された修正候補の良し悪しを振り分けることが可能である。

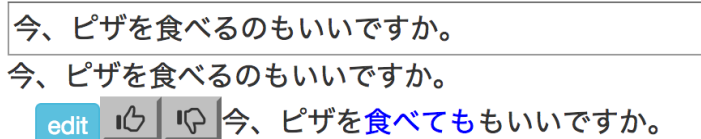


図4：サムズアップ・ダウンによるユーザーフィードバック機能

サムズアップ・ダウンの数を記録し、統計的にサムズダウンが多いものを削除していけば、誤った修正規則はシステムからなくなり、自然に規則のクリーンアップがなされる。

また、AI チュータから提案される修正は、必ずしも一つとは限らない。例えば、図5にあるように複数の修正文が提案されることがある。このような場合、サムズアップ・ダウンからの統計を利用し、サムズアップの投票の多いものから順に提示するという「修正文のランク付け」も可能である。

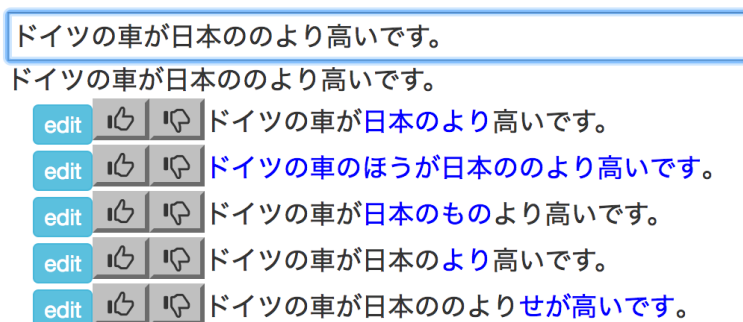


図5：複数の修正候補提示

一方、編集機能は、ユーザーに新しい修正文を入力してもらい、修正ルールを増やしていくという能動的なフィードバック方式である。現在のAI チュータは、図6にあるような編集機能を備え、ユーザーからの新しい修正規則が自動的にシステムの方に入っているようにしている。

今後、こうしたクラウドソーシングによる「人間の力」によっても、より洗練された修正規則の蓄積が可能になっていくことを期待する。

## 規則の編集

このケーキを**食べる**のいいですか。 → このケーキを**食べても**いいですか。

食べても		
Save	Remove	Cancel

図 6 : 複数の修正候補提示

## 5 おわりに

本稿では、Teacher Sourcing による誤用文<->修正文のコーパスデータから習得した誤用文修正知識を基にした日本語支援ツール、AI チュータの紹介をしたが、このプロジェクトは、「日本語教師とテクノロジーの協働作業」ということで、日本語教育と自然言語処理という二つの分野の架け橋にもなると期待される。日本語学習者が一体どのような誤りを犯すか、またそれをどのように修正したら日本語教育の観点から妥当なのか。これらの知識は、日本語教師のみがもっているエキスパート知識であり、AI チュータは、この日本語教師のエキスパート知識と自然言語処理技術との協働作業の賜物である。

今後の研究課題としては、AI チュータのカバレッジと精度の向上である。これらの問題に対しては、第 4 章でも触れたが、次の 3 つの観点から対処していきたい。

- 形態素解析を踏まえた修正規則の一般化
- 機械学習、ニューラルネットワークなどのテクニック利用
- ユーザーフィードバックによる規則の洗練

ただ、ここで大切なのは、データ確保の問題である。新しいテクニック、特に機械学習、ニューラルネットワークなどを交えることにより、コンテキストを踏まえた上での誤用修正ができるようになることを期待するが、こうしたテクニックには大量の訓練データが必要となり、今後いかに誤用<->修正のペアデータを確保していくかが問題解決の大切な鍵となる。その意味でも、一早く AI チュータをより多くの人に利用してもらい、本格的なクラウドソーシング

によるデータ収集を行いながら、システムの精度を向上させていくというエコ環境を整えていかなければいけない。

更に、今後 AI チュータを普及させるために不可欠なのは、AI チュータの日本語学習者の作文能力に対するインパクトの評価である。このためにも、将来的には AI チュータに誤用修正をさせるだけでなく、学習者にとって本当に有意義なスキヤフルディング(“scaffolding”)となるような情報をフィードバックとして出せるような支援ツールにさせていきたい。

---

<sup>1</sup> 研究における Teacher sourcing は国際交流基金ロサンゼルス文化センターからの助成を受けて行なったものであり、ここに感謝を記したい。

<sup>2</sup> 英語の場合、「have long hair」というように Have 動詞を使う。故に、ノンネイティブの修正に「ある」動詞が使われると仮定される。

<sup>3</sup> Abe (2016)は、この文は非文として取り扱っている。Abe は、この原因を数量詞と集合名詞の「家族」の共起に委ねている。

## 参考文献

Abe, Yasuaki. (2016) An Exploratory Essay on Error Analysis. In *Nanzan Studies on Japanese*

*Language and Culture Volume 16*, pp. 1--14.

大曾美恵子, 杉浦正利, 市川保子, 奥村学, 小森早江子, 白井英俊, 滝沢直宏, 外池俊幸

(1997) 「日本語学習者の作文コーパス: 電子化による共有資源化」, 『言語処理学会第3回年次大会論文集』, pp. 131--145.

大山浩美, 小町守, 松本裕治 (2012) 「日本語学習者の作文における誤用タグつきコーパスの

構築について—NAIST 誤用コーパスの開発—」, 『テキストアノテーションワークショップ予稿集』, pp. 1--8.

大山浩美, 小町守, 藤野拓也, 松本裕治 (2013) 「日本語学習者の作文におけるエラータイプの自動分類へ向けて」, 『第三回コーパス日本語ワークショップ』.

高橋哲朗, 相川孝子 (2016) 「日本語学習支援のためのコーパス構築システム」, 『第22回言語処理学会年次大会発表論文集』, pp. 689--692.

---

寺村秀夫 (1990) 『外国人学習者の日本語誤用例集』, 大阪大学: データベース版、国立国語研究所.

ドメニコ・ナガラ (1988) 『これは日本語か』 河出書房新社.

三上章 (1960) 『象は鼻が長いー日本文法入門』 くろしお出版

Mizumoto, Tomoya., Komachi, Mamoru., Nagata, Masaaki. and Yuji Matsumoto. (2011) Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners". In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP)*, pp. 147--155.

望月通子 (2012) 「日本語教育における学習者コーパスの構築と ICLEAJ」, 『関西大学外国語学部紀要』, 第 7 号, pp. 111--119.

吉永尚 (2012) 「テ形節における統語的考察」, 『園田学園女子大学論文集 第 46 号』, pp. 113--123.

李在鎬 (2012) 「コーパス分析に基づく構文研究」, 澤田治美 (編) 『ひつじ意味論講座第 2 巻 構文と意味』, pp. 241-265, ひつじ書房.

### ●さらに勉強したい人のために

1 黒橋 禎夫 (著) (2015) 『自然言語処理 (放送大学教材)』 放送大学教育振興会

基礎的なテキスト処理から構文や文脈の解析, そして機械学習などの応用など自然言語処理の概要が網羅的に説明されており, 自然言語処理の全体像をつかむために適している。

2 市川保子 (編著) (2010) 『日本語誤用辞典—外国人学習者の誤用から学ぶ日本語の意味用法と指導のポイント』 スリーエーネットワーク

外国人学習者が犯す誤用を日本語学的な観点から捉え、細かく解説している誤用例文の辞書。日本語教育の立場から、どのような指導を施したら効果的かという説明も細くなくされ、日本語教育に携わる者にとって、必見の書である。

3 益岡隆志、田窪行則 (1992) 『基礎日本語文法・改訂版』 くろしお出版

品詞や活用などの日本語文法における基礎が網羅的にまとめられている。本稿で述べた語句のパターンを理解するために有用である。