

# 英語学習者発話コーパスにおける誤り分析

## ー エラータグとその応用 ー

和泉絵美 (通信総合研究所)

井佐原均 (通信・放送機構/通信総合研究所)

### 1. はじめに

1999年より、日本人英語学習者の英語能力向上過程のモデル化および学習支援システムのプロトタイプ開発を目指し、日本人英語学習者発話コーパスを構築中である。学習者コーパスの特徴で、他の一般コーパスのそれと最も異なるのは、学習者による誤りが含まれているという点である。学習者にとって適切な教育法・支援を見出すための英語発話モデル構築において、学習者の誤り分析が有効であることは、英語教授法・第二言語習得の分野でも古くから言われてきたことである。本研究では、この学習者の発話書き起こしデータに含まれる誤りにタグを付加することによって誤り分析を実行することにした。しかし、学習者の誤りが、文法・語彙・語用・音韻など様々なレベルに亘ること、一つの誤りに対して複数の原因が推測されること、等の問題により、整合性の取れたエラータグ体系を構築するのは困難な作業となる。本稿では、まず伝統的な誤り分析の論理・方法について述べ、更に学習者コーパスにおける誤り分析の現状、本コーパスにおける誤り分析およびエラータグ体系について説明する。

### 2. 本コーパスの概要

対象データは、(株)アルクが実施するSST (Standard Speaking Test) というインタビュー形式のテストである。インタビューは、一被験者につき15分程度で、イラスト描写・ロールプレイ (店員と客のやりとり、友人同士の電話での会話など)・コマ割りされた漫画へのストーリー付けから成り立つ。これを収録し、書き起こし及びタグ付け・分析を行う予定である。発話データには、各被験者の英語能力の明示化として、SST独自の方法で判定された9段階のレベル

と、その他の主な英語試験の点数、判定結果等の情報が付加される。今年度すでに400人分の発話の収録・書き起こし及び基本タグ (フィラー・言い淀み・発話重複範囲の指定など) 付加を終了しており、発話の文法上の誤りに対してエラータグを付加している最中である。

### 3. 学習者コーパスにおける誤り分析

#### 3.1 誤り分析の目的

誤り分析の主な目的は、(a) 学習者がどの言語規則をどの順番で習得するか、という学習過程、(b) 学習者にとって目標言語のどの部分で正用法を導き出すのが困難か、を明示することである。これによって、学習者が習得過程のどの段階に位置し、次の段階に進むためには何を学習すべきかを指導することができる。[1] また、例えば本コーパスのように試験官 (教師) と受験者 (生徒) の間で多少でもインタラクションが存在するようなデータであると、それまで生徒が誤った形で使っていた箇所に対し、教師が正用の発話をする。生徒がそれを真似する (意識的であるか無意識であるかはいろいろだが。) ことによって、わずか15分のインタビューの間に生徒は正用法を (一時的にでも) 習得してしまうような例を見つけることもできた。

#### 3.2 誤り分析の手順

一般に、誤り分析の手順は、(あ) 誤り箇所の確認、(い) 誤りの分類、(う) 誤りの説明、(え) 誤りの評価、の4段階に分けられる。(あ) は学習者の発話のうち、どの単語・フレーズ・文構造・語順が誤っているか、という「誤りの場所」を指定する

ことである。ただし、すべての誤りに対してはつきりとした場所を限定できるとは限らない。形態素レベルの誤りの場所を指定するのは容易だが、例えば"**We have visited London last week.**"のような時制の誤りを含む文で、発話者は最初から"**We have visited …**"と言おうとしたのか、"**We visited …**"と言おうとして誤ったのか、によって誤り箇所は変化する。こういった文全体に関わる誤り (global errors) の範囲を限定するのは難しい。[2]

(い) は収集した誤りを体系化すべく、その誤りの言語範疇 (形態・統語・語彙など) や、表層構造が正用法とどう異なるかによって分類することである。

言語範疇の分類は、多くの伝統的な第二言語の教科書に見られるものと同様のものである。例えば、形態素レベルのものだと冠詞の誤用、動詞・形容詞・副詞の語形変化の誤用など。統語レベルだと主語・動詞の人称一致の誤用、語順の誤用など、という風に分類する。次に、表層構造の変化の分類は、目標言語の正用法と、学習者の誤用法が表層上どう異なるかという概念に基づいて、「添加・省略 (脱落)・置換・語順換え・それらの混合」の5つに分けられる。例えば、定冠詞"**the**"であるべきところを不定冠詞"**a**"に置き換えている場合は置換、冠詞の付け忘れは省略、などである。

次に、(う) の誤りの説明とは、何故その誤りが引き起こされたのか、誤りの原因を特定することである。誤りの原因として以下の5種類が一般的である。

- 母語の干渉によって引き起こされる誤り。  
(language transfer)
- 学習方略によって引き起こされる誤り  
(learning strategy-based errors)
- コミュニケーション方略によって引き起こされる誤り (communication strategy-based errors)
- 学習指導の方法によって引き起こされる誤り  
(transfer of training/induced error)
- 複数の要因によって引き起こされる誤り、その他原因が特定できない誤り

それぞれのタイプは更に下位分類できる。誤りの原因を明らかにすることは、学習者の学習ストラテジーや学習過程を推測するのに効果的である。

最後に、(え) 誤りの評価は、誤りを含む文の分かりやすさやネイティブらしさを評価することである。つまり、様々な誤りを、コミュニケーションの大きな妨げになるもの、それほど影響を及ぼさないものに分類し、誤りの重み付けをすることである。ここでは(あ)～(う)までのように、誤りを個別的に分類することとは違って、学習者の発話の全体を見渡すことが肝心となる。これは目標言語の母語話者によるチェックなどによって効果的に行われると考えられる。

### 3.3 誤り分析の問題点

誤り分析の意義、手法についてはいくつかの課題・問題点が挙げられる。一つの誤用を取ってみても、それが学習者の知識の欠如による純粋な「誤り」なのか、知識は持っているものの、それをうまく運用できず、思いがけず犯した「間違い」なのか判断するのは難しい。これは誤り分析の第一段階である誤り箇所の確認をする際の大きな問題となる。「間違い」は「誤り」と違って、一過性のものであるので、学習者の習得過程を包括的に見ることができない。[3]

他に、誤りの分類方法に明確で整合性の取れた基準がないこと、一つの誤りに対して複数の原因が考えられることが多い、等の問題もある。誤りの重み付けに関しても、非母語話者は文法項目などの局所的な誤りを重視するのに対して、母語話者は文の構造や表現方法といった全体的な誤りを重視する傾向にあるとされており、個人差が生じる可能性が高い。

### 3.4 学習者コーパスにおける誤り分析の現状

学習者コーパスにおいて積極的に誤り分析を行っている主な研究をいくつか紹介する。日本では、JEFLL (Japanese EFL Learner Corpus) コーパスが文法的誤りに8品詞別に分類したエラータグを付加している。語彙選択・コロケーション等の語用的誤りに関しては、COBUILD-direct のデータと比較して、学習者と母語話者の動詞"**make**"の用法の違いを分析している。海外では、学習者コーパスではないが、CHILDS (幼児の言語習得コーパス) において、CHAT

と呼ばれるフォーマットで、誤りを含む文の下にコメント形式で誤り箇所とそれに対する訂正文を記している。更に、エラーのレベルやタイプを分類してコード化したものを分析情報として付与している。この形式は、日本での大曽ら [4] による日本語学習者の作文コーパスにおいても、第二言語習得用にアレンジされて採用されている。また、ベルギーの Catholic University of Louvain の 15 カ国の英語学習者作文コーパス ICLE(International Corpus of Learner English) では、CHAT のようなコメント形式ではなく、POS タグなどと同じレベルでエラータグ体系が設計されている。[5]

#### 4. SST コーパスにおける誤り分析

本コーパスにおける誤り分析は、エラータグ付加を中心として進めているため、現時点ではより系統的に分類し易い文法的誤りのみを対象としている。付加する情報の分類は、先にも述べた伝統的な誤り分析の手法を用いている。つまり、まず誤り箇所(=タグ付加範囲)を特定し、あらかじめ分類された品詞や文法システム別のタイプ・表層構造上の変化のタイプを特定する。何故その箇所を誤りだと判断したのか、後々理解し易いように、誤りに対する訂正候補も明記する。誤りの原因についても情報付与したいところだが、整合性の取れた分類法が明確でないこと、一つの誤りに対して複数の原因が考えられることが多いことなどから、タグ付加作業者間での揺れが予想されるため現時点では保留としている。

エラータグのシンタックスは次の通りである。例えば、**"\*She has five childs."** という文に対しては、

```
ex) *She has five <er_ninf org="childs"
crt="children" mod="sub">childs</er_ninf>.
(注:er_ninf = 名詞の複数形への変化に関する誤り。
sub = 置換。)
```

という風にタグが付加される。まずその誤り箇所を開始タグと終了タグで囲む。タグの最初にはあらかじめ分類された品詞・文法システム別タイプのうちの一つを、**"er\_(品詞・文法システム)"**をというスタイルで明記する。タグ内の属性としては、**org(original)** に学習者が実際に発話した誤用を、**crt(correct)** に誤用に對

する訂正候補を、**mod(target modification taxonomy)** に表層構造上の変化のタイプを、それぞれ選択する。これは 3.2 で述べた手法によって分類されている。

また、一つの文に複数の誤りが含まれているような例においては、タグを入れ子で付加することになる。例えば、**"\*Apple is edible."** という文において、学習者は二つの誤りを犯したと解釈できる。一つは、**"apple"** に不定冠詞 **"an"** を付加し忘れていること。二つめは、一般的事象を述べるような表現では、名詞は複数形で用いなければならないが、単数形を用いていることである。これら二つの誤りについてタグ付加すると次のようになる

```
ex) *<er_nform org="an apple" crt="apples"
mod="sub"><er_at org="" crt="an" mod="omit">
</er_at> Apple</er_nform> <er_agr org="is"
crt="are" mod="sub">is</er_agr> edible.
(注:er_nform = 名詞の単複の誤り。er_at = 冠詞の誤り。omit = 省略。er_agr = 主語-動詞人称及び数一致に関する誤り。)
```

つまり、いきなり **"Apple"** を **"Apples"** に訂正するのではなく、まず冠詞の誤りとして **"an"** を補い、**"An apple"** に対する訂正として、**"Apples"** を補うことになる。この方法は、より詳細で用心深いと言えるが、結果的に実際には学習者が犯していない誤り(この例では、**er\_agr** の部分)にまでタグを付加することになり、ここまで詳細なタグ付加が必要かどうかは今後の検討課題である。

#### 5. 問題点・今後の課題

前項で述べたようなエラータグ体系に則って誤り分析を進めることにはいくつかの問題点が見られる。それらは、3.3 で述べたような誤り分析そのものの問題点に加えて、タグ付加作業は一部の事象を除いて自動化は難しく、人手での作業となり、膨大な時間と費用が掛かること、文構造全体または複数の文に亘る誤りへの体系的なタグ付加は難しいこと、などである。

また、コミュニケーションにおいて、時として文法よりも重要な役割を果たす語彙選択やコロケーションなどに関する誤りの分析方法をエラータグとは別に考える必要がある。これに関しては、現在のところ、

英語母語話者による正解コーパスを作成し、比較する案を検討中である。

また、誤り分析において重要な位置を占める誤りの原因の特定についても、主な事象を網羅しつつ、タグ付加作業者が混乱しないような形式の分類法を考える必要がある。誤りの原因は、母語の干渉によって引き起こされるものと目標言語内のものとに大別されるが、それらをより正確に区別するためにも、学習者コーパスの日本語訳 (back translation) コーパスを用意することも検討している。

## 6. まとめ

誤り分析は第二言語習得過程をモデル化し、適切な教授法を検討するための有効な手段である一方、その手法について多くの課題を抱えている。また、それをコーパスという膨大な言語資料の分析に適用するためには、様々なタイプの誤りの頑健な体系化が求められる。本プロジェクトでは、様々な誤りのレベル (形態素・統語) や誤りのタイプ (文法・語用) を、組織的に分類できるものは分類し、それが難しいものに関してはネイティブによるチェックによって対処したり、誤りの原因の特定に心理言語学的な分析を取り入れたりすることで、より多くの種類の誤りを分析対象とすることを目標とする。

また、データが学習者のレベルによって9段階に分けられているという本コーパスの特徴を生かし、学習者の誤りの種類・量を各レベル別に把握し、それぞれのレベルの学習者に最も適切な教授法を検討することも重要な課題のうちの一つである。

## 参考文献

- [1] ロッド・エリス著・金子朝子訳 1996: 第二言語習得序説。研究社出版。
- [2] James, C. 1998: *Errors in Language Learning and Use*, Addison Wesley Longman Ltd., London.
- [3] 小池生夫監修 1994: 第二言語習得研究に基づく最新の英語教育。大修館書店。

- [4] 大曾美恵子 他 1998: 日本語学習者の作文コーパス: 電子化による共有資源化。日本認知学会第15回大会。
- [5] Granger, S. (Ed.) 1998: *Learner English on Computer*, Addison Wesley Longman Ltd., London.