

翻訳学習者の学習過程におけるエラーの傾向の変化

山本 真佑花 田辺 希久子
神戸女学院大学

藤田 篤
情報通信研究機構

1 はじめに

翻訳の需要が高まる中、2015年5月にISO 17100が発行され[3]、翻訳者の資格要件として、翻訳に関する学位、翻訳以外の学位と2年以上の翻訳経験、5年以上の翻訳経験のいずれかを満たすことが求められるようになった。今後は、国内の大学・大学院における翻訳者の養成への期待も高まると考えられる。上記の要件が翻訳経験を含むことが示唆するように、学習者であっても、翻訳に関する教育を受ける段階から、実際に翻訳を行う経験を持つべきであろう。特に、学習者が翻訳を作成し、教員がそれを添削し、学習者がそれをふまえて翻訳を修正する、という実務翻訳のプロセスに沿った形式の翻訳実習が担う役割は大きい。

教員による翻訳の添削の際には、個々のエラーを一貫して判定すること、さらにはエラーの全体像を体系的に教授することが望ましい。これまで、そのような用途を想定してエラーの分類体系が作成されてきた[5, 2, 1, 7]。エラーを定量的に分析する道具があれば、学習者のエラーの傾向[6]やその通時的な変化[4]の観察が可能になる。さらには、観察された傾向を反映してカリキュラムを調整することも考えられる。

我々は、豊島ら[7]によるエラーの分類体系を実際の翻訳実習に導入し、エラーに関する体系的な説明、および学習者が作成した英日翻訳におけるエラーの指摘に活用している。その過程で蓄積した学習者のエラー事例に基づいて、学習過程におけるエラーの傾向の変化を観察し、学習によって減少しやすいエラーおよび減少しにくいエラーを明らかにした。

2 先行研究

Castagnoliら[2]は、MeLLANGEプロジェクトにおいて開発された3階層44カテゴリからなるエラーの分類体系[5]に基づいてエラータグ付きの学習者翻訳コーパスを構築し、頻出するエラーの種類や言語による差異を明らかにした。翻訳学習者向け共同翻訳プラットフォーム MNH-TT [1]では、この体系を単純化した2階層16カテゴリの体系を用いている。Toyoshimaら[6]は、この体系に基づいて学習者の英日翻訳におけるエラー709事例を分類し、次の3点を明らかにした。

- 全エラーのうち706事例が、16カテゴリのいずれかに分類できた。すなわち、元々は欧州言語対を

表1: 豊島ら[7]によるエラーの分類体系。

Lv 1	未完成
X4a	未翻訳
X6	曖昧さ未解消
Lv 2	誤訳
X7	用語の訳出エラー
X1	原文内容の欠落
X2	原文にない要素の付加
X3	原文内容の歪曲
Lv 3	目標言語の文法的または統語的な問題
X8	コロケーションのエラー
X10	前置詞や助詞のエラー
X11	活用のエラーや数・性などの不一致
X12	綴りエラー・誤変換
X13	句読法に関するエラー
X9	その他の文法的・統語的エラー
Lv 4	目標言語文書の質の問題
X16	結束性違反
X4b	直訳調
X15	表現のぎこちなさ
Lv 5	納品・公表に際しての問題
X14	レジスタ違反

対象として構築されたエラーの分類体系ではあるが、英日翻訳にも適用できることが確認された。

- 【X3 content-distortion】が最も頻出した。
- 学部生は、大学院生よりも顕著に多く、内容に関するエラーと文法に関するエラーを生じていた。一方、語彙に関するエラーの出現傾向については、両群に顕著な差は認められなかった。

ただし、いずれの研究においても、エラーの分類手続きは明確に定められておらず、分析の正確さに疑問が残る。豊島ら[7]はこれを指摘し、MNH-TTのエラー分類体系に基づいて、個々のエラーを一貫して分類するための判断基準を決定木と事例集の形で提案した。その際、学習者に対する指摘の優先度などを考慮して、表1に示すようにエラーの分類体系を再構築した。

学習者の翻訳能力の通時的な変化を観察した例として、Orozcoら[4]の取り組みが挙げられる。彼女らは、翻訳実習の開始前および終了後に、翻訳課題および質問紙調査を通じて、翻訳における文化差や言語固有の現象への対処方法、翻訳エラーの出現率、翻訳にかかわる一般知識の習得度を計測した。ただし、上記の3つの観点の各々について学習者の能力を1つの数値として表しているだけであり、例えばエラーの種類ごとの変化の多寡は明らかにしていない。

表 2: 起点言語文書とパッセージごとのワード数.

“Travellerspoint”の旅行ガイドの記事	T1	T2	T3
(t1) The Alps	192	195	184
(t2) Dubai	180	210	197
(t3) Taipei	168	167	209
“Amy’s column”のジャーナリズム記事	A1	A2	A3
(a1) Civil Rights	155	203	151
(a2) Free Speech vs. Surveillance in the Digital Age	178	155	205
(a3) Hiroshima and Nagasaki	175	216	164

3 翻訳学習者の通時的翻訳データの収集

我々は、翻訳学習者の学習過程におけるエラーの傾向の変化を観察するために、大学における実際の翻訳実習に豊島ら [7] によるエラーの分類体系を導入し、学習者による英日翻訳のデータを収集している。

本節では、ある大学において 2015 年 4 月～7 月に実施した翻訳実習について述べる。そこでは、次の 2 種類のテキストタイプの英語文書を題材とした¹。

- “Travellerspoint”²の旅行ガイドの記事
- “DEMOCRACY NOW!”³における “Amy’s Column”³のジャーナリズム記事

まず、各 Web サイトにおける同程度の難度の英語文書を 3 件ずつ選択した。そして、翻訳課題として手頃な分量(約 200 ワード)のパッセージに分割し、各文書の最初の 3 パッセージずつ(合計 18 パッセージ)を用いた。以下では、旅行ガイドの記事を分割して得たパッセージを起点言語文書の種類にかかわらず、T1, T2, T3, ジャーナリズム記事を分割して得たパッセージを同様に A1, A2, A3 と記す。パッセージごとのワード数を表 2 に示す。

今回の翻訳実習には学部生 27 名⁴が参加した。各受講生は、教員が用意した表 2 の文書の中から旅行ガイドの記事とジャーナリズム記事のそれぞれについて 1 件を選択し、期間の前半に T1, T2, T3 の 3 パッセージを、後半に A1, A2, A3 の 3 パッセージを翻訳した。各文書の翻訳を始める前に、パッセージごとに用語集の作成、背景知識に関する調査を各 1 名が担当し、クラス内での発表を通じて受講生全員で知識を共有した。対象とするパッセージの翻訳に必要な情報が欠けている場合は、教員から追加情報が与えられた。

¹翻訳実習では多様なテキストタイプに触れさせる必要があるためである。同じ大学で 2015 年の 9 月～2016 年 1 月にも今回と同種の題材を用いた翻訳実習を実施しており、今後、2 つの期間に共通する受講生の翻訳を分析する予定である。

²<http://www.travellerspoint.com/>

³http://www.democracynow.org/blog/category/weekly_column/

⁴受講生の大半は英語を専攻している。このうち 5 名は翻訳に関する授業を 1 年間受講した経験を、別の 3 名は半年間の入門講座を受講した経験を持つ。残りの 19 名は翻訳の初学者である。

受講生は、全 6 パッセージの翻訳課題の各々を、1 週目に下訳の作成、2 週目に指摘されたエラーの修正という手順で翻訳した。1 週目に提出された下訳に対して、教員とティーチング・アシスタント各 1 名が、豊島ら [7] のエラーの分類体系および決定木に従ってエラーを付与⁵、ダブルチェックし、各学習者に個別に返却した。各エラーは何らかの修正案を想定して分類されているが、学習者に対しては、エラー箇所と表 1 のエラーカテゴリのみが知らされた。学習者とその修正を唯一の正解と思い込んでしまう [5] ことを避けつつ、受講生にエラー分類を意識させ、教員と受講生の間にエラー分類に関する共通認識を生み出すことがねらいである。

この翻訳実習を通じて最終的に、学習者のエラー事例 1,707 件を得た。

4 エラー傾向の変化の分析

上述の翻訳実習で得られたエラー事例に基づいて、学習者のエラー傾向の変化を観察した。翻訳実習においては、起点言語文書を短いパッセージに分割して翻訳させたが、今回の分析では、前半の 3 パッセージ (T1, T2, T3) と後半の 3 パッセージ (A1, A2, A3) を各々まとめ、それらの 2 期間における原文 100 語あたりのエラーの出現回数(以下、**相対エラー数**)を比較した。

4.1 学習者全体の傾向

前半の 3 パッセージ分と後半の 3 パッセージ分の各々における相対エラー数(学習者ごとの値のマクロ平均と標準偏差)を表 3 に示す。起点言語文書と目標言語文書の両方に関係する Lv 1 および Lv 2 の相対エラー数の合計は約 5%増加した(3.93 → 4.11)。先行研究 [6, 7] と同様に最も頻度が高かった【X3 原文内容の歪曲】は約 30%増加した(2.20 → 2.91)。テキストタイプごとの難度の影響もあるだろうが、英語文書の読解能力が容易には向上しないことを示唆しているとも解釈できる。一方、全体で 2 番目に多かった【X7 用語の訳出エラー】は半分以下に減少した(1.14 → 0.53)。用語集を参照することで容易に確認・修正できることに加え、今回の翻訳実習では、豊島ら [7] の決定木に従って比較的高い優先度でこのエラーを指摘したことにも効果があったと考えられる。

目標言語の表現に関する Lv 3 および Lv 4 の相対エラー数の合計は約 30%減少した(1.52 → 1.08)。目標言語は母語であるため、エラーの体系に対する理解度や

⁵同一パッセージ内で同じ用語の訳出エラー (X7) や同じ種類のレジスタ違反 (X14) を生じていた場合は、初出の 1 度のみをカウントおよびフィードバックした。

表 3: 原文 100 語あたりの相対エラー数: 学習者ごとの値のマクロ平均 (M) と標準偏差 ($S.D.$).

エラー	前半		後半	
	M	$S.D.$	M	$S.D.$
Lv 1 X4a	0.03	0.07	0.03	0.07
X6	0.00	0.00	0.00	0.00
Lv 2 X7	1.14	0.79	0.53	0.39
X1	0.47	0.42	0.49	0.31
X2	0.09	0.13	0.16	0.19
X3	2.20	0.95	2.91	1.03
Lv 3 X8	0.11	0.11	0.13	0.18
X10	0.19	0.24	0.18	0.27
X11	0.07	0.09	0.12	0.15
X12	0.15	0.16	0.11	0.11
X13	0.09	0.16	0.17	0.21
X9	0.22	0.34	0.03	0.08
Lv 4 X16	0.07	0.11	0.03	0.07
X4b	0.53	0.53	0.24	0.22
X15	0.10	0.18	0.08	0.13
Lv 5 X14	0.30	0.25	0.45	0.31
合計	5.76	2.17	5.65	1.84

翻訳に対する意識が高まるにつれて、課題の提出前に文法的・統語的な問題を排除したり、ごちない表現を修正したりすることが能動的にできるようになったということだろう。特に、この範囲で最も頻出していた【X4b 直訳調】のエラーが半分以下に減少した (0.53 → 0.24) ことに注目したい。豊島ら [7] が報告している非学習者かつ非教員によるエラー分類実験では、分類誤りの中で【X3 原文内容の歪曲】と【X4b 直訳調】の混同が最も多かった。これに対して、今回の学習者にとってのこれらの 2 種類のエラーの改善しやすさは、顕著に異なっていた。

Lv 5 の【X14 レジスタ違反】は約 50% 増加した (0.30 → 0.45)。後半のジャーナリズム記事においては、引用や中略が多用されるが、馴染みがない学習者が多く、それらの表現に関する指定に違反しがちであった。

4.2 学習者ごとのエラーの傾向

上述の 4 種類のエラーカテゴリについて、相対エラー数が増加・減少した学習者数、および前半と後半の相対エラー数の相関係数を表 4 に示す。全エラーの中で、【X3 原文内容の歪曲】が最も高い相関 ($r = 0.691$) を示した。翻訳実習において繰り返し指摘しても、相対エラー数の傾向には大きな変化がなかった。一方、学習者全体で相対エラー数が半減した【X7 用語の訳出エラー】および【X4b 直訳調】については、多少増加している者も多少おり、相関係数も高くはなかった ($r = 0.240, r = 0.250$)。言い換えれば、後半のジャーナリズム記事における相対エラー数は、前半の旅行ガイドの記事における相対エラー数に必ずしも依存していない。

表 4: 相対エラー数が増加・減少した学習者数および前半と後半の相対エラー数の相関係数。

エラー	増加	変化なし	減少	相関係数
X3	22	0	5	0.691
X4b	7	2	18	0.250
X7	4	0	23	0.240
X14	20	2	5	0.221

4.3 学習者のクラスタ分析

相対エラー数の分布に基づいて学習者のクラスタ分析も行った。まず、学習者間の距離を全エラーの相対エラー数のユークリッド距離で、クラスタ間の距離をウォード法で定義し、前半の旅行ガイドの記事と後半のジャーナリズム記事の各々について樹形図を構成した。距離の閾値を 5 としたとき、旅行ガイドの記事については 5 クラスタ、ジャーナリズム記事については 3 クラスタが得られた。この時の学習者の分布を表 5 に示す。赤く塗られたクラスタ C_1, C'_1 が、各期間で最も相対エラー数が少なかったクラスタを表す。あわせて、相対エラー数に基づく学習者のグラフを図 1 に示す。これらは、学習者の全ての対をユークリッド距離の昇順に走査することで最初に得られた (つまり最小の) 連結グラフである。ノードは学習者を表し、表 5 に対応する色を塗ってある。

全エラーの中で最も出現割合が高く、かつ学習者ごとの標準偏差も最大である【X3 原文内容の歪曲】が、クラスタの形成においても最も大きな影響力を持っていた。前半の C_1, C_2, C_3 、および後半の C'_1, C'_2, C'_3 の違いは、概ねこのエラーの多寡のみで説明できる。

前半のグラフには、これとは別の方向に 2 つのクラスタ C_4, C_5 が形成された。これらは、【X7 用語の訳出エラー】や【X4b 直訳調】を特によく生じた学習者で構成されている。4.1 節では、全体的な傾向としてこれら 2 種類のエラーの半減を観察したが、個々の学習者に共通する (表 4 の相関係数が高くなるような) 傾向ではなく、表 6 に示すように、 C_4 と C_5 に所属する 8 名がこれらのエラーを、他の 19 名よりも顕著に大きく減少させた影響であることが明らかになった。一部の学習者に固有の相対エラー数の減少を反映する形でエラーの傾向の多様性も減少し、図 1 が示すように学習者のグラフがより直線的になった。

5 おわりに

我々は、豊島ら [7] によるエラーの分類体系を実際の翻訳実習に導入し、エラーに関する体系的な説明、および学習者が作成した英日翻訳におけるエラーの指摘に活用している。その過程で蓄積した学習者のエラー

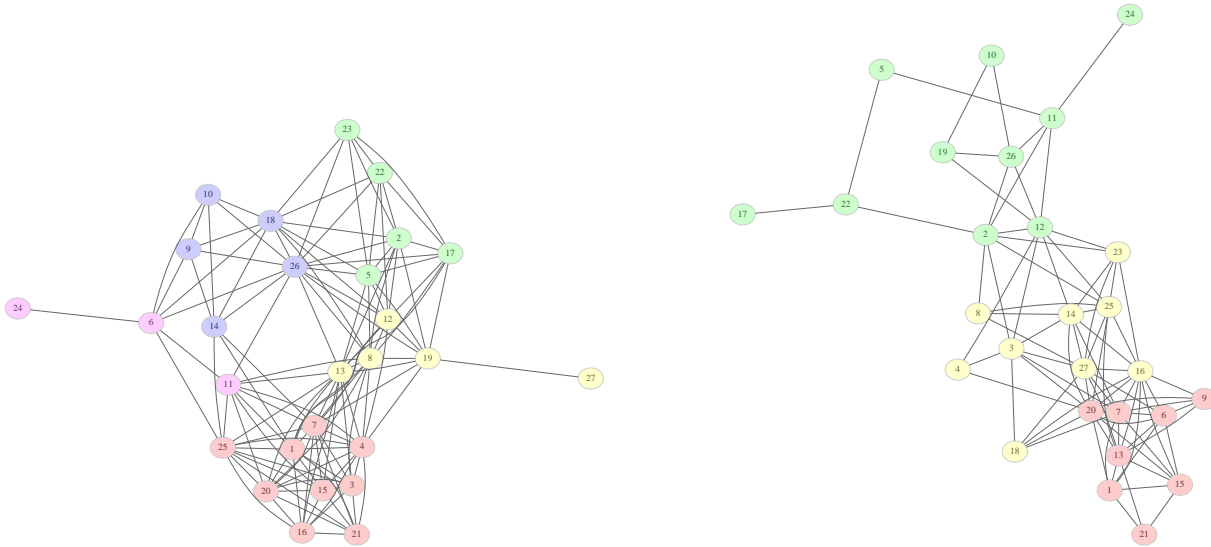


図 1: 相対エラー数の類似性に基づく学習者のグラフ (左: 前半, 右: 後半).

表 5: 各クラスターに所属する学習者.

前半 \ 後半	C'_1 (8 名)	C'_2 (9 名)	C'_3 (10 名)
C_1 (9 名)	1, 7, 15, 20, 21	3, 4, 16, 25	-
C_2 (5 名)	13	8, 27	12, 19
C_3 (5 名)	-	23	2, 5, 17, 22
C_4 (3 名)	6	-	11, 24
C_5 (5 名)	9	14, 18	10, 26

表 6: クラスター間の相対エラー数の比較.

前半の クラスター	X7			X4b		
	前半	後半	差分	前半	後半	差分
C_1	0.58	0.25	-0.33	0.23	0.15	-0.08
C_2	0.61	0.53	-0.08	0.60	0.34	-0.26
C_3	1.14	1.02	-0.12	0.32	0.38	+0.06
C_4	1.93	0.43	-1.50	1.69	0.36	-1.33
C_5	2.21	0.58	-1.63	0.48	0.11	-0.37

事例に基づいて、学習過程におけるエラーの傾向の変化を観察し、次の3点を明らかにした。

- 【X7 用語の訳出エラー】および【X4b 直訳調】のエラーは減らすことができる。特に、初期の状態ではこれらのエラーを比較的多く生じる学習者でも、指摘を繰り返すことによって、他の者と同程度まで減らすことができる。
- 【X3 原文内容の歪曲】は最も頻出するが、それだけ指摘されても全体的に減少しにくい。
- 【X14 レジスタ違反】は馴染みのないテキストタイプを扱う際に増える。

エラーの分類体系や決定木は、翻訳実習において教員と学習者の間でエラーの判定基準や深刻さなどについての共通認識を生み出すためのメタ言語となる。学習者は、エラーのカテゴリに応じて適切な情報を参照し、エラーを回避・修正することができるようになる。

教員にとっても指導の効率化・円滑化が可能になる。今回の翻訳実習を担当した教員からも、手応えがあるとの声が寄せられている。

今後は、テキストタイプを固定した分析を予定している。エラー付与作業の効率化や自習の活性化のために、エラーの自動検出についても検討を進めている。学習者に対する適応的な翻訳実習の観点から、学習者の特性を早期に診断する技術についても検討したい。

謝辞: 本研究の一部は科研費基盤研究(A) (課題番号: 25240051, 代表: 影浦峽) の支援を受けた。

参考文献

- [1] B. Babych, A. Hartley, K. Kageura, M. Thomas, and M. Utiyama. MNH-TT: a collaborative platform for translator training. In *Proceedings of Translating and the Computer* 34, 2012.
- [2] S. Castagnoli, D. Ciobanu, K. Kunz, N. Kübler, and A. Volanschi. Designing a learner translator corpus for training purpose. In *Proceedings of the 7th International Conference on Teaching and Language Corpora*, 2006.
- [3] ISO 17100:2015. Translation services: Requirements for translation services, 2015.
- [4] M. Orozco and A. Hurtado Albir. Measuring translation competence acquisition. *Translators' Journal*, Vol. 47, No. 3, pp. 375–402, 2002.
- [5] A. Secară. Translation evaluation: A state of the art survey. In *Proceeding of the eCoLoRe/MeLLANGE Workshop*, pp. 39–44, 2005.
- [6] C. Toyoshima, K. Tanabe, A. Hartley, and K. Kageura. Error categories in English to Japanese translations. In *Proceedings of the 21st Annual Meeting of the Association for Natural Language Processing*, pp. 1076–1079, 2015.
- [7] 豊島知穂, 田辺希久子, 藤田篤, 影浦峽. 翻訳教育での利用を意識した翻訳エラー分類体系の再構築. 言語処理学会第22回年次大会発表論文集, 2016. (in this proceedings).