

翻訳プロセスへのコーパス分析とジャンル概念導入の有効性

宮永正治 (大阪大学情報科学研究科)

照井雅子 (大阪大学言語文化研究科)

三崎敦子 (サイマル・アカデミー)

野口ジュディー (武庫川女子大学薬学部)

概要：翻訳は単語を別の言語のものとただ置き換えるだけではない複雑な作業であることはよく知られているが、このプロセスの効率を高める試みを行った。日英翻訳者養成ワークショップに応用言語学の ESP (English for specific purposes) アプローチを取り入れ、参加者にジャンル分析やコーパスツールを利用して翻訳タスクに取り組んでもらった。ジャンルを意識して翻訳に取り組むと、ソーステキストとターゲットテキストのレトリック構造が明らかになり、表現選択の判断基準も分かりやすくなる。十人の参加者は課題文書の分析を行い、言語的特徴について議論した。議論前の第一稿と議論後の第二稿の比較から、コーパス分析とジャンル概念導入の有効性が見られた。

1. はじめに

英語と日本語では同じジャンルに属するテキストであっても、レトリック構造が異なっていることが多い。そして、レトリック構造が異なっている場合、ソーステキストにある文をそのまま訳しても、ターゲット言語におけるジャンルテキストとしては通用しない。Newmark (1988) は、最も自然に、また効率よく翻訳できるのは外国語から母国語への場合であると言っている。では、日本語話者が英語に翻訳する場合、どのような手法によれば、効率よく高品質の翻訳ができるのであろうか。Swales (1990) は、move と step という単位を用いて、科学分野における研究論文の introduction の部分のレトリック構造を明らかにした。我々は、このようなジャンル分析を基礎とする ESP (English for specific purposes) アプローチ、すなわち、ソーステキストおよびターゲットテキストの Purpose, Audience, Information, Language feature (PAIL) を Observe, Classify, Hypothesize, Apply (OCHA) すること (野口 2003) の有効性を、日英翻訳ワークショップの参加者の翻訳テキストをとおして検証することにした。

2. 日英翻訳ワークショップ

2-1. 概要

今回報告する日英翻訳ワークショップでは、ESP の手法および言語学的ツールを使用した。本ワークショップは、企業のホームページを用いて、実践的な翻訳手法の習得、ソーステキストとターゲットテキスト間のギャップを解決すること、そして言語学的ツールの使用法の習得を目的として 2006 年 7 月に始まった。講師は ESP の専門家であり翻訳者でもある英語母語話者。参加者は計 10 名で、フリーランス翻訳者、人材派遣会社勤務、銀行管理職、証券トレーダー、公務員

など職業は様々ではあるが、皆、同じ翻訳クラスのエ受講者である（受講期間は1年から数年と異なる）。翻訳作業は複雑なプロセスであるために、本ワークショップでは、各参加者がリソースを収集・活用し、語法のパターンを検証、そして各翻訳原稿に対応する語彙文法的パターンを特定することを狙いとした。ソーステキストおよびターゲットテキストのジャンルについては、ジャンル分析およびコーパス言語学的ツールを用いて分析した。分析には、課題に用いたテキストと同じ各ジャンルでコーパスを構築して使用した。参加者たち自身が発案、企画し、運営していることも本ワークショップの特徴であり、講師からの指示だけではなく、参加者による課題テキストの選定から、コーパス用データの収集、およびその他の議論については、そのほとんどをメーリングリストを通じて行った。

2-2. 内容

- ・ソーステキスト：企業のプレスリリース（酒造メーカーの化粧品およびサプリメント新発売についてのプレスリリース）

(1) 第1日目までの課題

1) コーパスの構築

全員で化粧品およびサプリメントについての英文プレスリリースを収集し、コーパスを構築した。

2) 翻訳第1稿の作成

各自で第1稿を作成した。

(2) 第1日目

翻訳中の問題点およびその解決法、そしてコーパス検索結果について参加者が発表し、講師からは move analysis の手法が紹介された。

(3) 第2日目までの課題

1) Move analysis

コーパス用に収集したテキストについて、各自で move analysis を実施した。

2) 翻訳第2稿の作成

第1日目の議論に基づいて翻訳第2稿を作成した。第2稿は第2日目までに講師が校閲した。

(3) 第2日目

レトリック構造、文法、語彙、および表記法について、校閲済みの課題を用いて議論した。

3. 分析方法

参加者の第1稿におけるエラー、第2稿におけるエラー、および第1稿から第2稿で変更した点を error taxonomy を用いて分類した。第1稿から第2稿で変更した点については、エラー以外で修正した点についても分類の対象としたが、エラーを修正したところ、およびエラー以外を修正したところのいずれについても、修正する前のエラーおよび（エラーでない場合は）問題点と思われる点について分類した。Table 1 に分類項目を記す。尚、分類に当たっては、McCoy ら (1996) の error taxonomy を参考にし、McCoy らの分類にないものは独自に項目を設けた。また、文法 (G)、

語彙 (L)、レトリック構造 (R)、表記法 (T)、または文脈 (S) のレベルについても分類した。さらに文脈 (S) のレベルについては、第 1 稿で原文にない情報を追加していたか (Information added)、情報の解釈が間違っていたか (Information mistaken)、または、ジャンル上、直訳すべきでないところを直訳していたか (Direct translation) についても分類した。

上記のように分類した項目のうち、ジャンルテキストの作成上重要と考えられるレトリック構造 (R)、および文脈上変更した点 (S) のうち Omitted paraphrase (Direct translation および Information missing) について、各参加者がどれだけ修正を行ったかをみた。

Table 1 Error taxonomy

Error	GLRTS	Semantic factor	Error	GLRTS	Semantic factor
Collocation	G		New sentence	R	
Extra determiner	G		Redundancy	R	
Inappropriate determiner	G		Sentence separation	R	
Inappropriate preposition	G		Sentence structure	R	
Noun formation	G		Omitted paraphrase	S	Direct translation
Omitted determiner	G		Omitted paraphrase	S	Information missing
Verb usage	G		Paraphrase	S	Information mistaken
Mispelling	L		Paraphrase	S	Information added
Word choice	L		Capitalization	T	
Extra determiner (Heading)	R		Description	T	
Inappropriate tense	R		Double-byte character	T	
Inappropriate tense	R		Extra capitalization	T	
Itemization	R		Inappropriate punctuation	T	
Itemization	R		Number	T	
Layout	R		Omitted punctuation	T	
Name	R		Punctuation	T	
New paragraph	R		Spacing	T	

4. 結果と考察

Table 2 に、第 1 稿におけるエラー数の平均値、第 2 稿における修正の数の平均値、および第 2 稿における修正の数の平均値と第 1 稿におけるエラー数の平均値の割合を示す。第 2 稿における修正の数の平均値が第 1 稿におけるエラー数の平均値より高いのは、第 1 稿のエラーでないところを第 2 稿において改善したことを示す。

Table 2 No. of errors in the first translation and improvements in the second translation on average

Error class	Average No. of errors in first translation	Average No. of improvements in second translation	Improvements/Errors
R	1.7	8.1	4.8
S (Omitted paraphrase: Direct translation)	2.5	1	0.4
S (Omitted paraphrase: Information missing)	0.9	0.1	0.1
G	8	1.1	0.1
L	4	9.6	2.4
S (Paraphrase: Information added)	1.3	0.9	0.7
S (Paraphrase: Information mistaken)	1.4	0.9	0.6
T	22.2	2.6	0.1

ジャンルテキストへの翻訳上重要と考えられるレトリック構造 (R) についての修正は他の分類項目よりも高く、またエラーでないところを改善した数も高い。また、ジャンルテキストに翻訳する際に、ソーステキストの表現を paraphrase する必要がある場合があるが、その点 (Omitted paraphrase) については、あまり高くなく、各翻訳者のジャンルテキストに対する認識が必要とされることを示している。

参考文献

- McCoy, K. F., Pennington, C. A. & Suri, L. Z. (1996). *English error correction: A syntactic user model based on principled mal-rule scoring*. In Proceedings of the Fifth International Conference on User Modeling
- Newmark, P. (1988). *A textbook of translation*. New York: Prentice-Hall.
- 野口ジュディー サイバーメディア・フォーラム(Cybermedia Forum) No. 4 大阪大学サイバーメディアセンター Teaching ESP Writing: OCHA in a CALL Class 40 頁～48 頁 (2003).
- Swales, J. M. (1990). *Genre Analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.