

下訳から修正訳への訳文修正要因の分析

阿辺川 武

影浦 峯

東京大学大学院 教育学研究科

{abekawa,kyo}@p.u-tokyo.ac.jp

1 はじめに

翻訳の世界では、下訳から最終訳までを 1 人の翻訳者が担当する場合と、下訳と最終訳を別々の人が担当する場合が存在する。後者の場合は、下訳を翻訳経験の浅い翻訳者が担当し、下訳を修正する形で最終的な訳をベテランの翻訳者が担当する形式が多い。我々は現在、翻訳者支援システムの開発を行っており [3]、その一環として一般書の翻訳における下訳とその修正訳のデータを収集している。特に下訳は、最終的に本として出版される修正訳とは異なり、世の中に流通しないことから貴重なデータといえる。そのため下訳と修正訳を対象にした研究は、現在、筆者の知る範囲では見あたらない。下訳と修正訳を比較することで、翻訳過程がどのように進行していくか、また翻訳経験の差により訳文の質がどれだけ異なるのかなど、対訳コーパスの利用だけでは得ることの難しい、翻訳に関する様々な知識が獲得できると期待される。最終的には、修正差分のデータをわかりやすい知識にまとめ、初心者翻訳者の学習用に、そして修正候補箇所の自動検出に向けた知識として利用することを目標としている。

本稿では翻訳における下訳とその修正訳を見比べ、どのような要因で修正されているかを分析・整理し、コーパスにタグ付けをおこなった結果を報告する。我々は最初に、修正要因を記述するためのタグセットを構築した。タグには修正を生じさせる要因を記述した「理由」と、どのような修正を施したかを記述した「操作」がある。さらに理由はどのような「状態」にあるかと、どのようにしたいかという「目的」から構成され、また「操作」においては、主に文法的に記述することを目的に「操作目的 → 文法操作 → プリミティブ操作」という 3 階層構造を定義した。

2 修正要因の分類

修正パターンの発見の前に、翻訳文の修正過程では修正者がどのような意図を持って修正を施すかについての我々の考えを整理しておく。修正を記述するにおいては、修正を生じさせる要因を説明した「理由」と、

それにともない実際にどのような修正を行なったかという「操作」に大別することができる。

2.1 理由

理由はさらに、これから修正を加えようとする訳文がどのような「状態」にあるかと、その訳文に対してどのような修正を施したいかという「目的」に分けられる。我々は修正の理由を次の表 1 のように分類したが、実際には、もう少し具体的な言葉で記述できる項目もある。

表 1: 修正の理由

状態 (~である)	目的 (~させる)
1. 誤訳である	誤訳を正す
2. 文意がうまく伝わっていない	文意を正しく伝える
3. 訳文がぎこちない	なめらかな訳文にする
4. 編集の方針と合っていない	編集の方針に合わせる
5. 修正翻訳者の嗜好に合わない	好みの表現にする
6. その他 (上記に分類できない)	

理由についてはあきらかな誤訳を除き、修正者が修正時に頭の中で考えていたことであり、我々が、英文と下訳、修正訳の 3 文対を後から見ても正確な理由をうかがい知ることはできない。

2.2 操作

理由にともない、修正者が実際に下訳に対して修正を行なった現象のことを操作とする。下訳と修正訳の 2 文対を形態素解析し、形態素列を比較することにより表層上の修正差分は獲得できる。しかし、単純に表層上の修正差分だけでは、何を換えようとして修正したのか、修正された要素同士に関係はあるのかなど、より大きな視点から修正要因を把握することは難しい。そこで我々は、どのような目的で修正操作を行なったかを記述する「操作目的」、その中で「文を受動態から能動態に直す」といった文法的にどのような操作がなされたかを記述する「文法操作」、そして個々の形態素がどのように変化したかを記述する「プリミティブ操作」と、操作の記述に 3 階層を設け、修正操作全体をわかりやすく把握できるようにした。

なお誤訳の修正に関しては、翻訳文の読み易さ以前の問題であるため、修正操作の記述は行わず、理由において誤訳であることを指摘するに留める。

操作目的

修正操作の目的を記述したもので、上記の理由1つに対して、1つ以上の操作が結びつけられる。我々がまとめた操作目的の一覧を表2に掲載する。数字は後で述べる今回タグ付けしたデータ中の頻度である。

表 2: 操作目的

流暢化 (読点の挿入, 接続表現の挿入など)	95
より適切な表現へ	33
構文の流暢化	32
不足情報の付加 (主に内容語)	26
提題化	13
固有表現の書き換え (「アムコ->アムコ社」など)	13
スタイルの変更	10
冗長情報の削除 (主に内容語)	9
関係の明確化	8
話し手と出来事の時間軸の変更 (テンス, アスペクト)	7
容易・軽い表現化 (「明瞭に->はっきりと」など)	7
主観表現の変更 (モダリティー)	8
照応詞の具体名詞化	2
ゼロ代名詞付加	2
重い表現化 (「輸入と輸出->輸出入」など)	1
節のバランス (長い修飾節の分割など)	1
冗長表現の削除 (主に機能語)	1

文法操作

文法操作は、修正操作を文法的に記述したものであり、文法用語を用いて記述する。現時点では、「動詞の態の変換」や「連体修飾から連用修飾へ」といった、修正された要素が表す機能を記述する方針としている。またテンスやモダリティーなどについては、その要素が表す意味も併記している。一方で、単に品詞の操作だけで機能が記述できない操作もあり、この辺りは若干方針から逸脱するタグ定義となっている¹。現時点は、試行錯誤の段階であり、今後事例を増やしながらか一貫性のあるタグ定義を整理していく予定である。

プリミティブ操作

一番細かい単位の操作で形態素ごとの操作を記述したもので、「挿入」「削除」「書き換え」「移動」の4種類がある。表層の操作なので客観性が高く、将来的には自動処理によりタグを付与できると考えている。

表 3: 文法操作の例

削除: 名詞, 挿入: 助詞, 移動: 動詞句
テンスの変更: 過去->現在
格の変換: 「によって」->「が」
書き換え: 助詞: 助詞->助詞
動詞の態の変換: 受動->能動
アスペクトの削除: 完了
係り先変更: 連用修飾句->連体修飾句
モダリティーの変更: 推定->推定
用言の活用変換: 連用テ接続+テ->連用中止
品詞変更: 副詞->名詞句

表 4: プリミティブ操作の例

挿入 「、」、挿入 「すべて」
削除 「には」、削除 「という」
書き換え 「に」->「を」
書き換え 「ながら」->「つつ」
移動 「馬鹿げた説明を」<->「定期的に」

3 データ

本研究で用いているデータは、石油の枯渇に関する英語の一般向け書籍 [5] と、それに対する日本語訳 [6] である。文献 [4] による分類でいえば、「中間的テキスト」という位置付けであり、「文学テキスト」ほど原文の持つ文章のリズムやうねりなどを再現する必要はなく、「科学・特許・公式文書」ほど定型表現を用いた訳し方が存在するわけでもなく、翻訳の文体は訳者による裁量の余地がある程度ある。

現在手元にある翻訳データは、翻訳経験の浅い3人が分担して下訳を担当し、それに対して経験12年の翻訳者が1人で修正を施したものである。データの基本的な統計データを表5に掲載する。なお、簡単なDPマッチングによる下訳と修正訳の形態素単位の一致率は約78.9%であった²。

表 5: データの統計

	英文	下訳	修正訳
文数	4,414	4,628	4,648
単語 or 形態素数	81,207	127,721	132,823
(一文あたり平均)	18.4	27.6	28.6

翻訳者ごと、および下訳と修正訳でどのくらい品質に差があるかを調べるためにnグラム言語モデル³を用いてパープレキシティーを計算した結果を表6に示す。

下訳を翻訳する翻訳者によってパープレキシティーに差があるのは、翻訳経験や文体のくせにより個人差があるからであると思われる。そして修正訳では一貫し

²本研究では、形態素解析器に Mecab を使用した。
http://mecab.sourceforge.net/

³コーパスは Web から抽出した日本語文約 4,000 万文を使用し、n グラムの接続数は n=4 とした。

¹修正の単位の問題もあり、どこまでも1つの修正として扱うか、句と節にはどのような違いがあるかなど明確な区別をしていない。

表 6: n グラムモデルによるパブレキシティー

	全文	翻訳者 A	翻訳者 B	翻訳者 C
下訳	66.2	66.8	57.3	70.5
修正訳	59.0	58.5	52.7	62.9

て言語モデルとしてはよくなっているが、下訳からの向上度はどの翻訳者による翻訳でも同程度であり、修正は下訳に大きく引き摺られることがわかる。

今回は、上記のデータの中から、翻訳者による翻訳者 A による下訳 60 文についてタグを付与した。タグ付与は、翻訳経験を有する言語学者 1 名により行われた。タグを付与したデータの例を表 10 に掲載する。

4 分析

現時点でタグ付けされたデータについて、頻度の集計を中心に分析を行なった。

4.1 各タグの頻度

理由についての内訳を表 7 に掲載する。60 文という少なさとはいえ修正箇所が数百あるなかで誤訳による修正が 10 件しかないということは、経験の浅い翻訳者でも原文の意味の把握はほぼ正しくできているといえよう。

表 7: 理由の内訳

1. 誤訳である	10
2. 文意がうまく伝わっていない	14
3. 訳文がごちない	164
4. 編集の方針と合っていない	24
5. 修正翻訳者の嗜好に合わない	14
6. その他 (上記に分類できない)	1

操作目的の内訳については、すでに表 2 において記したが、流暢化が一番多い。これは流暢化が範疇とする操作目的が多岐に渡るためであることが一因であるが、一般に、翻訳の大きなポイントが対象言語としての自然さにあることも原因である⁴。

4.2 原文との近さ

各修正操作について、文構造や意味を総合的に判断して、下訳と修正訳とでどちらが原文の英語に近いかを主観で記述してもらった。集計結果は表 8 のようになり、下訳の方が原文に近い表現となっていることが多いとわかるが、修正訳の方が近いことも少からずあり、下訳の翻訳者が原文にとらわれない独自の表現で翻訳したものを、修正者により修正されている現象は興味深い。

表 8: 原文との近さ

下訳の方が近い	90
修正訳の方が近い	25
どちらとも言えない	122

修正訳の方が原文に近い翻訳例を以下に挙げる。

原文:	... the community of people who have worked in the others
下訳:	その他の<>油田に入り込んで働いてきた</>多くの人々
修正訳:	ほかの<>油田で働いてきた</>多くの人々

4.3 読点

読点に関しては、機能までを詳細に記述した。文献 [2] で論じられているように読点の機能は多岐に渡り、研究者により分類はそれぞれ異なっているが、本データでは出現した事例に対し独自に機能を記述した。

表 9: 読点の機能

挿入:主題明確化	23
挿入:分割	6
挿入:文修飾 その他	6
挿入:直後の修飾を回避	4
挿入:文修飾 連用テ形後接	4
挿入:文修飾 副詞後接	3
挿入:並列句明確化	2
読点の削除	2

5 おわりに

本稿では、翻訳において一般には流通することの少ない下訳と修正訳のデータに対して、どのような要因で修正されるのか、そしてどのような修正がなされているかについて、実際のデータに対しタグ付けを行い、作成されたデータを元に分析を行なった。元来、文を修正するという作業は、修正者の翻訳経験や言葉の捉え方など客観的でない要素が大きく、それを修正された結果のみからタグを付与していく作業は困難を要することが今回のデータ作成において判明した。

今回付与したタグは暫定的なものであり、今後の課題として、タグ付与の基準の作成や用語の統一をしていくことでデータの客観性を増すとともに、ごちない翻訳文の自動検出といった機械処理へと向けたデータのありかたを検討することが挙げられる。

付記

本研究の一部は、日本学術振興会科学研究費補助金基盤 (A) 「翻訳者を支援するオンライン多言語レファレンス・ツールの構築」(課題番号 17200018)、および財団法人博報児童教育振興会の支援を得て行われた。

⁴それでも、連語的表現についてもとからその言語で書かれたテキストと翻訳テキストのあいだには違いがあることが指摘されている [1]。

表 10: タグが付与されたデータの例

<p>原文: It is important to realize that the early toppers are not advocates or agitators by choice.</p> <p>下訳: 早期ピーク論者がどちらかと言えば提言者でも煽動者もないということを認識することは重要である。</p> <p>修正訳: 大切なのは、早期ピーク論を主張する研究者が、なにか意図をもって提言しているわけでも、煽動しようとしているわけでもないことを理解することである。</p>	
<p>修正 1 早期ピーク論者がどちらかと言えば提言者でも煽動者もないということを<2>認識することは</2><1>重要である</1>。</p> <p style="text-align: center;">↓</p> <p><1>大切なのは、</1>早期ピーク論を主張する研究者が、なにか意図をもって提言しているわけでも、煽動しようとしているわけでもないことを<2>理解することである</2>。</p> <p>理由: 3. 訳文がぎこちない 操作目的: 提題化 文法操作: <1>品詞変更:動詞句->名詞句 <2>品詞変更:名詞句->動詞句 プリミティブ: <1>移動「重要である」 <1>削除「である」 <1>書き換え「重要」->「大切」 <1>挿入「なのは」 <1>挿入「,」 <2>書き換え「認識」->「理解」 <2>削除「は」 <2>挿入「である」</p>	<p>修正 3 <>どちらかと言えば</></p> <p style="text-align: center;">↓</p> <p><>なにか意図をもって</></p> <p>理由: 1. 誤訳である</p>
<p>修正 2 <>早期ピーク論者</></p> <p style="text-align: center;">↓</p> <p><>早期ピーク論を主張する研究者</></p> <p>理由: 3. 訳文がぎこちない 操作目的: 不足情報の付加 文法操作: 挿入:動詞 書き換え:名詞:名詞->名詞 プリミティブ: 挿入「を」, 挿入「主張する」 書き換え「者」->「研究者」</p>	<p>修正 4 <1>提言者でも</1><2>煽動者も</2>ないということ</p> <p style="text-align: center;">↓</p> <p><1>提言しているわけでも</1><3>, </3><2>煽動しようとしているわけでも</2>ないこと</p> <p>理由: 3. 訳文がぎこちない 操作目的: 構文の流暢化 不足情報の付加 文法操作: <1><2>品詞変更:サ変名詞+名詞 ->サ変名詞+スル <3>挿入:読点:並列句明確化 プリミティブ: <1><2>削除「者」 <1>書き換え「提言」->「提言する」 <1>挿入「ている」 <1><2>挿入「わけ」 <2>書き換え「煽動」->「煽動する」 <2>挿入「ようと」 <2>書き換え「も」->「でも」 <3>挿入「,」</p>
	<p>修正 5 ない<>という</>こと</p> <p style="text-align: center;">↓</p> <p>ない<></>こと</p> <p>理由: 3. 訳文がぎこちない 操作目的: 流暢化 文法操作: 削除:助詞相当句 プリミティブ: 削除「という」</p>

参考文献

- [1] Carmen Dayrell. A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics*, Vol. 12, No. 3, pp. 375-414, 2007.
- [2] 原誠. 読点使用のルールに関する覚え書. 文体論研究, Vol. 52, pp. 1-39, 2006.
- [3] 影浦峽, 佐藤理史, 竹内孔一, 宇津呂武仁, 辻慶太, 小山照夫. 翻訳者支援のための言語レファレンス・ツール高度化方針. 言語処理学会第 12 回年次大会, pp. 707-710, 2006.
- [4] 影浦峽, 阿辺川武. 翻訳者の類型と翻訳作業の諸相. 言語処理学会第 13 回年次大会, pp. 919-922, 2007.
- [5] Jeremy Leggett. *Half Gone: oil, gas, hot air, and the global energy crisis*. Portobello Books Ltd, 2005.
- [6] ジェレミーレゲット (原著), 益岡賢, 楠田泰子, 植田那美, リックタナカ (訳). ピーク・オイル・パニック-迫る石油危機と代替エネルギーの可能性. 作品社, 2006.