

英日翻訳における下訳と修正訳の語彙的/文体的特徴の分析

村山 遼† 影浦 峯†

† 東京大学大学院 教育学研究科

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

1 はじめに

ブログに代表される各種情報発信メディアの普及にともない、職業翻訳者ではない人間が個人的な興味関心にもとづいて時事・社会問題に関する翻訳をおこない、ウェブ上で公開するという事例は増えつつある。その際には、どうしたら自然で読みやすい訳文をつくることができるかという問題が、翻訳の質を左右する大きなポイントとなる。特に、英日翻訳のように原言語と対象言語の構文構造が大きく異なる言語間の翻訳では、原言語の構造を過剰に意識するあまり、原文の意味解釈においては正しいものの、日本語として不自然で読みにくい訳文をつくってしまう、という状況が、翻訳初心者の場合にはしばしば起こりがちである。

このようなぎこちない翻訳は、翻訳者養成のための教科書においても敬遠されることが多い。たとえば安西 [4] は「翻訳」を「直訳」「英文和訳」と対比させながら、受動態や関係代名詞など、英語に頻出する構文を類型化し、それらを自然な日本語に翻訳するためのテクニックをまとめている。しかし、大半の教科書では、原文と自然な訳文との対応関係が強調される一方で、そもそもどのような日本語表現がぎこちないのか、という点についての言及はなされていない。したがって、ぎこちなさを解消するために訳文のどこをどう直したらよいか、という判断は、依然として熟練翻訳者の手に委ねられている。

本研究の目的は、このような翻訳者の暗黙知を明らかにすることにある。訳文のぎこちなさに気づき、自然で読みやすい表現に修正するという翻訳にとってのプロセス [5] を、立ち戻って日本語表現内部の問題としてとらえ直すことで、従来の教科書とは違った角度から、翻訳者（とりわけ翻訳初心者）を支援していきたい。

本稿ではその第一歩として、翻訳の下訳と修正訳のテキストを比較分析し、「ぎこちなさ」を事例に即してきめ細かく分析していくための手がかりと方向性を定めることを試みる。なお、本研究の最終的な目標は、与えられた訳文からぎこちない日本語表現を検出し、代替候補の訳文を提案する

システムの構築である。

2 ギこちなさという問題

広義にとらえた翻訳の「質」を左右する主要因として、誤訳の有無、クオリティ・コントロール（文字・表記の統一、専門用語の一貫性等、比較的厳密に統制可能な要素の統制の度合い） [6]、訳文の読みやすさ/ぎこちなさ、がある。誤訳が、少なくともその典型的なものについては、原文に照らして明確に判定でき、また、クオリティ・コントロールの概念も比較的明確にされているのに対し、ぎこちなさはほとんど定義上、厳密化に抗する概念であり、具体的な翻訳表現を前にしたときにも、(1) それでぎこちなさかどうかの判断には誤訳やクオリティ・コントロールよりも判断者間の揺れが一般に多く、また、(2) そもそも何がぎこちなさを構成する表現上の要因であるかも、少数のパターンを除いては十分明確になっているとはいえない。また、「ぎこちない表現でありうるような構成」を持つ表現クラスがどのようなインスタンスにおいて実際にぎこちないと判断されるかはさらにわかっていない。

こうした曖昧さが生み出す問題を回避するために、阿辺川ら [3] は、下訳データ、完成訳データに機械翻訳データを加えてテキストのマクロな統計的特徴を比較し、さらに機械学習を行うことによってぎこちなさの定義や個別的な特徴の分析的析出を回避したかたちでぎこちなさを検出することを試みているが、十分に目標を到達しているとは言えない。他方で、翻訳関係の教科書がぎこちなさの規定を例示と学習者の理解力とに任せてしまっているのは「はじめに」で述べた通りである。このような状況を考えるならば、現段階で、改めてぎこちなさをその一般的な概念規定とデータに即した分析的な要因析出との両面から明らかにすることが重要であると思われる。本稿で、下訳と完成訳データの全体的な特徴を見るのは、そのためである。

3 データ

本研究で用いるデータは、石油の枯渇に関する一般向け書籍の翻訳過程で生まれた下訳と修正訳

である。下訳は翻訳経験の浅い3人(2~5年)の分担によって、また修正訳は経験12年の翻訳者1人によって作成された。この下訳に施された修正の約8割が、誤訳や編集方針よりもむしろそのぎこちなさに起因している[2]。したがって、下訳の語彙的/文体的特徴をぎこちなさな日本語表現に近似することはおおむね妥当である。テキストの基本的なデータを表1に掲載する。

表1から、十分顕著な変化と考えられるのは、総単語数と総文節数の増加である。一般に構造の異なる言語間の翻訳を行うとき、対象言語のなめらかで自然な表現を作ろうとすると、表現は、おそらくは原言語の構造に影響をより強く受けていると推測される下訳の表現よりも、長くなりがちであることが伺える。文数、一文あたり平均単語数と一文あたり平均文節数もそれぞれ修正訳では増えているが、顕著な傾向と言えるほどではない。

表1 基本的データ

	下訳	修正訳
文数	4,493	4,497
単語数	130,574	135,573
一文あたり平均	29.06	30.15
文節数	46,948	47,468
一文あたり平均	10.45	10.56

4 分析

4.1 品詞の全体的傾向

ぎこちなさを特徴づける要素を析出するために、下訳と修正訳について、品詞レベルでの数量的特徴、機能語の特徴、機能語を語彙化したnグラム(n=2, 3)などを観察した。今回は、そのうちもっとも基本的なところについて分析する。

表2からは、極めて興味深い全体的な傾向が観察される。第一に、のべ数で見ると、すべての品詞で、修正訳の方が出現頻度が高いことである。これは、全体に表現が長くなっていることに対応している。

第二に、記号と連体詞を除くすべての品詞で、修正訳の方が、異なり語数が少ないこと(したがって平均使用頻度は修正訳の方が高く、「語彙の豊富さ」は下訳の方が高いこと)である。これについては、暫定的に、以下のように考えることができよう。すなわち、一般に原言語の表現と構造に引きずられると考えられる下訳の段階では、原言語

のレトリックや表現法に対応して多様な語が使われるが、日本語として改めて読み直す際にそれにより文意・文脈が取りにくいと判断され、それらが一定程度平板化される可能性(用語の統一といったクオリティ・コントロールが下訳では十分になされていないことも影響している可能性がある)。ただし、これを検証するためには、より多くの多様な下訳・修正訳データが必要である。

表2 品詞別の出現頻度

	下訳		修正訳	
	のべ	異なり	のべ	異なり
名詞	51,196	5,459	52,021	5,275
助詞	35,703	110	35,674	98
動詞	15,671	1,352	16,286	1,333
記号	12,558	32	15,723	37
助動詞	9,832	19	9,880	18
副詞	2,064	301	2,129	287
連体詞	1,219	32	1,364	35
形容詞	1,135	149	1,216	134
接続詞	592	57	688	55

4.2 助詞

助詞の出現頻度を調査したところ、格助詞が占める傾向は下訳・修正訳ともに変わらず、のべ語数の増加にもなって「を」「が」「て」などは若干の増加をみせている(表3)。一方で「の」「に」「は」などの出現頻度は低下しており、特に「の」の減少が著しい。本稿には掲載していないが、その大きな理由は「<名詞>の<名詞>の<名詞>」といった3つ以上の体言を「の」で接続する表現が軒並み削除されていることによる。具体的な事例を挙げるならば、「無色無臭のガスの充満」を「無色無臭のガスが充満して」に書き換えるなど、句を節に変換する操作や、「石油の帝国の指導者」を「石油帝国の指導者」として複合名詞化する操作が修正訳では頻発している。また、連語「による」「によって」の頻度も低下しており、これは阿辺川[1]が取り上げた受動態から能動態への変換を裏づけるものとなっている。「に関する」「に対する」といった表現も、上位15件には入らないものの減少をみせている。この点に関しては考察で言及する。

連語の中でも減少の著しかった「という」に関しては、後続する単語列について調査した(表4)。

形式名詞を他の名詞と別個に扱い集計をおこなったところ、「～ということ」「～というもの」「～というの」といった形式名詞をとまなう表現が修正訳において敬遠されていることが明らかになった。

表 3 助詞の出現頻度 (上位 15 件)

下訳		修正訳	
の	6,864	の	6,493
に	4,930	に	4,866
は	4,741	は	4,719
を	3,830	を	4,044
が	3,336	が	3,503
て	2,631	て	2,810
と	1,687	と	1,856
で	1,581	で	1,599
も	1,113	も	1,141
から	586	から	567
か	477	か	472
という	443	ば	345
ば	334	まで	316
や	331	や	309
まで	295	という	296

4.3 接続詞

接続詞によって文同士の論理関係を明示することは、流暢で読みやすい日本語文をつくる 1 つの手法である。下訳と修正訳を比較すると、順接を示す「そして」は大きな変化がみられない (表 5)。「けれども」「しかし」「だが」「ところが」といった逆接表現は若干の増加をみせつつ、「けれども」の増加、「しかし」の減少など、その分布が変化している。ただし、この現象が本当にぎこちなさの解消に貢献しているのか、あるいは単に修正者の好みによるものかは、現段階では判断できない。原因を表す「なぜなら」も減少している。並列や添加を表す「また」は増加したが、「または」については減少している。さらに、「すなわち」のように、直前の文を説明したり、文中で用いて表現の言い換える役割を果たす接続詞が増加していることも興味深い。

4.4 読点

文のリズムを整えたり、統語構造を明確化する際に、読点は大きな役割を果たすと考えられる。表 6 で明らかのように、修正訳では大半の場合において読点が新たに挿入されている。特に、助詞「は」

表 4 「という」に後続する単語列 (上位 15 件)

下訳		修正訳	
という-名詞-を	40	という-名詞-を	39
という-こと-だ	33	という-名詞-の	21
という-こと-を	30	という-名詞-だ	19
という-名詞-が	26	という-名詞-が	19
という-の-は	25	という-名詞-名詞	19
という-名詞-の	22	という-名詞-は	18
という-名詞-名詞	22	という-名詞-で	12
という-こと-は	19	という-名詞-に	11
という-名詞-だ	18	という-の-は	11
という-名詞-に	18	という-読点-名詞	10
という-もの-だ	16	という-もの-だ	8
という-こと-に	15	という-こと-に	7
という-名詞-は	15	という-こと-は	7
という-名詞-で	12	という-点-だ	7
という-わけ-だ	11	という-こと-だ	7

の直後に読点を挿入するケースは、修正訳において格段に増加している。また「を」の直後に読点を挿入しているのも修正訳独自の特徴である。これらはたとえば入れ子状の複文の係り受け関係を明示する役割と果たしていると思われる。

5 考察

以上の結果から、ぎこちなさの一端を掴むことができた。これらのぎこちなさが生じる理由については、原文である英語の文法構造の観点から、いくつかの仮説的な説明が可能である。

たとえば安西 [4] は “the Norman’s conquest of England” を「ノルマン人のイギリスの征服 (は)」は「ノルマン人がイギリスを征服したこと (が)」と書き改めるべきと主張しているが、同等の操作の結果が、今回分析したデータにおける「の」減少となって現れている。しかし「拙訳の典型」であるとした連続する「の」以外にも「ノルマン人によるイギリスの征服 (は)」という訳が修正される可能性も考えられ、これが「による」(あるいは「に関する」「に対する」といった、名詞を接続する表現全般の減少につながっていると思われる。むしろ “by” を「による」と直訳した表現が「拙訳の典型」として修正され、「による」が削除される場合もあるが、これらのパターンの分類は、現段階では困難である。今後データを原文と比較することで明らかにしていきたい。

表 5 接続詞の出現頻度 (上位 15 件)

下訳		修正訳	
そして	84	けれども	99
けれども	64	そして	93
しかし	62	また	57
たとえば	35	たとえば	38
しかも	27	しかし	38
また	25	すなわち	32
だが	20	しかも	28
または	18	ところが	21
いっぽう	15	あるいは	20
もつとも	15	いっぽう	18
すなわち	14	ただ	17
なぜなら	14	しかしながら	16
ところが	13	および	15
および	13	つまり	15
あるいは	12	もつとも	14

表 6 読点の挿入される直前の単語 (上位 15 件)

下訳		修正訳	
は	1,323	は	2,347
が	364	が	607
に	344	に	515
で	279	で	391
し	204	も	243
ば	194	ば	218
も	162	と	214
と	158	し	202
て	119	て	185
あり	93	を	147
か	90	ため	96
ず	79	か	95
れ	75	けれども	91
ため	75	あり	85
によって	61	ず	83

「ということ」「というもの」については名詞節を構成する that 構文をどう訳すかに関わっていると考えられる。

6 おわりに

本稿では、下訳データと修正訳データをマクロな観点から対比分析することにより、「ぎこちなさ」に関連している可能性のあるいくつかの現象と要素を指摘した。本稿の分析はぎこちなさの表現の析出、ぎこちなさという概念の明確化、さらには、ぎこちなさという表現でありうるクラスがどのようなインスタンスでぎこちなさのものとなるかの同定に向けた第一歩にすぎない。分析対象としたデータの規模と範囲も極めて限られているため、現在のところ、本格的な分析のための手がかりを得たという程度にすぎない。

今後は、

- (1) 下訳・修正訳データの規模・範囲の拡大および参照データ（翻訳ではなくもともと日本語で書かれたデータ等）の利用
- (2) 今回、明確になった要素の文脈を含めた特徴の質的な析出と「ぎこちなさ」との関係の分析および「ぎこちなさ表現」クラスのリストアップ
- (3) 既往の翻訳指南書と(2)の比較整理

などの面から分析を拡張・深化させていく予定で

ある。

参考文献

- 1) 阿辺川武, 影浦峯. (2007). 下訳と修正訳を用いた訳文修正パターンの発見. 言語処理学会第 13 回年次大会. 919-922.
- 2) 阿辺川武, 影浦峯. (2008). 下訳から修正訳への訳文修正要因の分析. 言語処理学会第 14 回年次大会. 253-256.
- 3) 阿辺川武, 影浦峯. (2008). 機械翻訳を利用したぎこちなさの翻訳文の検出への試み. 言語処理学会第 14 回年次大会. 325-328.
- 4) 安西徹雄. (1995). 『英文翻訳術』筑摩書房.
- 5) Mossop, B. (2007). *Revising and Editing for Translators. 2nd edition.* St. Jerome Publishing.
- 6) Thelen, M. (2009). Quality Management for Translation. *CIUTI-Forum 2008: Enhancing Translation Quality: Ways, Means, Methods.* 195-212.