下訳と修正訳を用いた訳文修正パターンの発見

阿辺川 武　　影浦 峡
東京大学大学院 教育学研究科
abekawa@p.u-tokyo.ac.jp　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なっており [4]，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．

また，従来の言い換えの研究 [2] では，語あるいは言語表現の内包的な意味が同一であれば言い換えであると仮定しているが，その内包的な意味をいかにして捉えればよいかは明らかではない．そして，言い換えではニュアンスあるいは言外の意味の変化もともなうが，どのような言い換えがこれらに影響するかも明確にはなっていない．一方，下訳から修正訳への過程を言い換えであると考えると，表現すべき意味は英文により示されており，言い換えるべき目的が明確に存在する．したがって，あきらかな誤訳を修正する場合を除き，獲得できる言い換えの多くは，英文の意味内容を正しく表現するためのニュアンスや言外の意味の変化を含有していると思われる．

本研究では，以上のことを念頭に置きながら，下訳から修正訳に至る過程でどのような要因で修正されるのか，その際どのような修正操作が施されるのかを解明することを目的とする．最終的には，翻訳初心者が作成したぎこちない翻訳文に対し修正候補が提示できる機能の開発を目標とする．本稿では，研究の第一段階として下訳と修正訳を用いて，典型的な修正パターンを発見する手法，およびその結果について報告する．

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．

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)	

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 文対を比較することにより，表層上の修正差分として獲得できるものである．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．操作の種類には，動詞の態を変え主語と目的語の格を変換するといった複数の箇所を一度に修正するものもあれば，ある単語の表現を 1 つ置き換えるといった単独で行なわれる修正もある．

そこで我々は，操作の記述に階層を設け,「文を受動態から能動態に直す」という上位階層の操作と，その中で発生する「ヲ格の格要素をガ格へ」や「動詞を受動形から能動形へ」といった個々の文節単位の修正からなる下位階層の操作を定義する．本稿では，この下位階層の操作のことをプリミティブな操作と呼ぶ．

本研究の目的は，翻訳文の修正過程を解析するにあたり，典型的な修正パターンを発見することにある．本稿ではその手法として，まず下訳と修正訳の比較からプリミティブな操作を抽出し，次に共起する複数のプリミティブな操作をまとめ上げ，大きな修正パターンとして獲得する手続きをとる．本研究ではプリミティブな修正操作をパターンとして集積しやすい構文的な操作に限定する．表 1 で挙げた主な修正理由のうちでは主に 3,4,5 の理由に該当する修正操作を獲得することになる．

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[1] と，それに対する日本語訳[2] である．文献 [5] による分類でいえば,「中間的テクスト」という位置付けであり,「文学テクスト」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大きい．

現在手元にある翻訳データは，翻訳経験が 2 年および 5 年の翻訳者による下訳 (だいたい半々ずつ) と，それに対して経験 12 年の翻訳者が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 2 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[3]．

表 2: データの統計

	英文	下訳	修正訳
文数	4,587	4,629	4,648
単語 or 形態素数	92,300	127,838	132,989
(一文あたり平均)	20.1	27.6	28.6
文節数	–	47,158	47,823
(一文あたり平均)	–	10.2	10.3

4 プリミティブな修正操作の獲得

4.1 アライメント

最初に前処理として，下訳と修正訳について，DP マッチングを用いて文単位の対応をとる．1 対多対応の場合は，長い文に合わせ 1 文として扱う．この DP マッチングの過程で形態素単位でのアライメントも取得できるが，順序が入れ換っている場合を考慮していない．そこで順序の入れ換えを考慮したアライメントを取得するために，本研究では本来対訳コーパスのアライメントで使用される GIZA++[7] を使用する．GIZA++ による下訳と修正訳のアライメントの例を図 1 に示す．対訳コーパスと異なり，同じ言語の間でも説明の付加や削除が頻繁におこなわれているため，正しくアライメントができないことがある．

4.2 プリミティブな修正操作

本稿でいうプリミティブな操作とは下訳と修正訳のアライメントの結果から形態素単位で異なる箇所を操作として表現したものである．主に構文的な操作をプリミティブな修正操作として獲得したいため，誤訳を正す際に生じる語義の選択誤りの修正や，翻訳者の嗜好による単なる語彙の置換といった構文操作と関連しない操作は対象としない．さらに，構文操作が多く含まれる格要素と用言を含む文節に限定して修正操作を獲得する．

格要素では，格の変換や格要素から被修飾名詞への転置などであり，用言では態の変換，活用形の変換，サ変名詞の名詞形から動詞形などが挙げられる．基本的に，お互いの操作間は包含関係にならないように定義している．下訳，修正訳の対からプリミティブな修正操作を獲得した例を表 3 に掲載する[4]．また,「仕事をする」「掘削を行なう」といった「サ変名詞+を+機能動詞」の場合や「することができる」など複数の文節に分かれてしまう表現は，係り受け解析の後処理として 1 文節にする．

表 3: プリミティブな修正操作の例

下訳 修正訳	彼によって 彼が	油田の 油田を	掘削が行なわれた 掘削したこと
操作	によって → が	の → を	削除-機能動詞 削除-られる 挿入-こと:形式

4.3 結果

頻出する修正操作に対し，動詞の活用形についてを表 4 に，名詞の格変換についてを表 5 に示す．なお「削除-サ変名詞形」とは，格要素としてのサ変名詞が用言に転じたものであり,「挿入-サ変名詞形」とはその逆である．助詞の変換については，例えば「が → を」の場合，ガ格の削除，ヲ格の挿入としてカウントしている．

表 4 より，下訳から修正訳にかけて受動態が減少しているだとか,「(∼して) しまう」「(∼して) もらう・くれる」のような訳者の主観が含まれてしまうような表現が減少しているなど全体的な傾向が見てとれる．

助詞については，特に IPA の品詞体系でいう「助詞・格助詞・連語」の減少数が著しい．これには大きく 2 つの要因が考えられ，1 つは助詞・格助詞・連語の使

[1]Jeremy Leggett,*"Half Gone: oil, gas, hot air, and the global energy crisis"*,Portobello Books Ltd,2005

[2]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.

[3]本研究では，形態素解析器に Mecab を使用した. http://mecab.sourceforge.net/

[4]この他にも動詞の自他,「取れる」「取らす」などの可能形，使役形などの情報は文献 [3] を参照している．

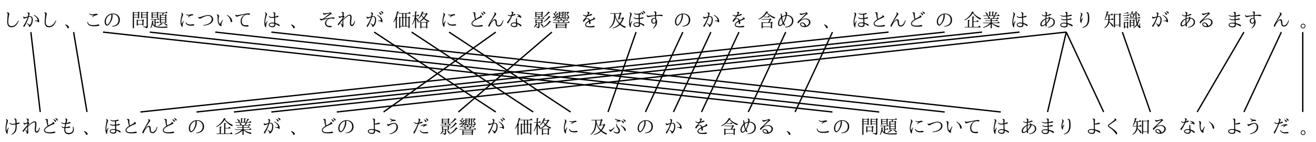

図 1: GIZA++によるアライメントの例

用がテクストの要請する表現と適合しない，もう 1 つは助詞・格助詞・連語の使用が直訳調の翻訳に頻出し，自然な日本語を求める修正者により軒並削除されているということである．助詞・格助詞・連語についての考察は別途，機を改めて報告したい．

表 4: 動詞の活用形

活用	削除	挿入	活用	削除	挿入
だ	379	291	こと	41	36
て	269	358	しまう	58	14
た	247	306	よう	54	51
られる	224	122	させる	48	39
いる	197	267	くる	31	46
の	188	150	する	31	43
ある	128	267	なる	22	26
ない	102	139	しれる	20	21
う	93	106	くれる	18	5
できる	76	82	ます	17	4
その他					
サ変名詞形	208	184	を+機能動詞	34	30

表 5: 助詞の変換

格助詞	削除	挿入	助詞・連語	削除	挿入
に	476	384	によって	103	13
が	387	502	に対して	44	10
の	366	204	について	34	36
を	293	421	として	32	22
で	203	193	による	30	10
と	75	101	において	27	3
から	47	41	に関して	26	3
へ	20	21	を通じて	14	1
その他助詞			に対する	14	8
は	398	371	に関する	13	6
副助詞	99	123	という	13	31
接続助詞	18	24	にとって	10	17
その他			より	3	10
連体修飾	88	30	により	2	29

5　修正パターンの発見

複数のプリミティブな修正操作間の共起をとらえ，修正パターンを獲得する．共起の範囲を広くとると，得られたパターンが示していることの把握が困難になる恐れがあるため，今回は範囲を用言とそれに係る格要素間に限定する．本研究では CaboCha[5]を用いて係り受け解析を行なっている．範囲内のすべてのプリミティブな修正操作の組み合わせについて対数尤度比を計算し，共起関係の強い対を表 6 に示す．

次に表 6 で得られた共起対について，これらを 1 つの修正操作と考え，同じように対数尤度比を計算し，3 つ組の修正パターンを獲得した．表 7 にその結果を示す．

5.1　考察

表 6 の結果から，典型的な修正パターンはだいたい 3 つに分類できる．(1) 受動・能動の態の変換，(2) サ変名詞の名詞形と動詞形の変換，(3) 格要素と被修飾名詞の変換，である．これら構文操作は従来の言い換えの研究で議論されてきたものであり，本手法で獲得した結果は新鮮味があるものではない．

翻訳初心者向けに書かれた翻訳指南書 [1, 6] などでは，英日翻訳において，このような英語構文があるときは日本語ではこの構文に変換すればよいというように，構文を対応づけて説明することが多い．今回得られた修正パターンは，英文と下訳という言語の差こそあれ，その指南書にあるパターンがほとんどであった．つまり経験の浅い翻訳者は，英文の構文をそのまま日本語でも同じ構文として翻訳し，ベテラン翻訳者は結果的に翻訳指南書と同じ翻訳方法に基づいて修正訳を構築していることがわかる．

6　おわりに

本稿では，最初に下訳から修正訳に至る過程で，どのような理由をもとに修正が成されるかについての分類を行ない，修正の理由にともなって施される修正操作を形態素単位のプリミティブな操作の集合として定義した．そして，実際の下訳と修正訳のデータからプリミティブな修正操作を獲得し，共起関係の強い操作対を獲得することで，下訳に対する典型的な修正パターンとして獲得する実験を行なった．本稿で利用した修正データは，単一の修正者からなり，特定のテクストに依存するもので，得られた結果の一般性を論ずることはできない．今後は多くの修正者によるデータを収集していくとともに，本データの特性を活かし，英文を含めた 3 文対を利用し，どのような要因で今回得られたような修正が施されるのかを明かにしていくこと

[5] http://chasen.org/~taku/software/cabocha/

表 6: 2 つの修正操作の共起

LLR	f(a,b)	f(a)	f(b)	操作 a	操作 b	書き下し表現
146.2	28	35	224	によって→が	削除-られる	A によって B される →A が B する
105.2	34	90	224	が→を	削除-られる	A が B される →A を B する
91.7	34	115	208	の→が	削除-サ変名詞形	A の B→A が B する
90.9	26	61	208	の→を	削除-サ変名詞形	A の B→A を B する
36.3	15	68	168	に→を	自動詞→他動詞	A に B する →A を C する
33.3	11	34	184	を→の	挿入-サ変名詞形	A を B する →A の B
31.8	12	65	122	を→が	挿入-られる	A を B する →A が B られる
27.2	11	61	130	の→を	挿入-こと:形式	A の B→A を B すること
21.8	7	28	130	連→が	挿入-こと:形式	B する A→A が B すること
21.1	4	5	208	の→連	削除-サ変名詞形	A の B→B する A
19.2	8	33	184	が→の	挿入-サ変名詞形	A が B する →A の B
18.6	9	65	123	を→が	他動詞→自動詞	A を B する →A が C する
15.9	5	61	39	の→を	挿入-させる	A の B→A を B させる
14.9	4	9	184	に→へ	挿入-サ変名詞形	A に B する →A への B
14.2	5	15	208	による→が	削除-サ変名詞形	A による B→A が B する
13.3	5	65	48	を→が	削除-させる	A を B させる →A が B する
13.3	5	90	35	が→を	によって→が	A が B によって →B が A を
11.1	4	121	20	は→が	挿入-とき:形式	A は B する →A が B するとき
11.0	5	19	224	で→は	削除-られる	A で B される →A は B する
11.0	5	33	123	が→に	他動詞→自動詞	A が B する →A に C する
10.2	8	121	93	は→が	削除-う	A は B だろう →A が B する

表 7: 3 つの修正操作の共起

LLR	f(a,b,c)	操作 a	操作 b	操作 c	書き下し表現
31.2	7	は→が	削除-う	削除-だ	A は B だろう →A が B する
25.2	5	が→を	によって→が	削除-られる	A が B によって C される →B が A を C する
21.2	7	の→が	削除-サ変名詞形	他動詞→自動詞	A の B→A が C する
18.9	10	の→が	削除-サ変名詞形	挿入-た	A の B→A が B した
16.3	6	の→が	削除-サ変名詞形	挿入-られる	A の B→A が B される
16.0	6	の→が	削除-サ変名詞形	挿入-こと:形式	A の B→A が B すること
4.6	6	の→が	削除-サ変名詞形	挿入-て	A の B→A が B して

が課題である．

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) の支援を得て行われた．

参考文献

[1] 安西徹雄. 英文翻訳術. 筑摩書房, 1995.

[2] 乾健太郎, 藤田篤. 言い換え技術に関する研究動向. 自然言語処理, Vol. 11, No. 5, pp. 151–198, 2004.

[3] 情報処理振興事業協会技術センター. 計算機用日本語基本動詞辞書 IPAL(Basic Verbs)–解説編–. 1987.

[4] 影浦峡, 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.

[5] 影浦峡, 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, 2007. 発表予定.

[6] 中村保男. 英和翻訳の原理・技法. 日外アソシエーツ, 2003.

[7] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, Vol. 29, No. 1, pp. 19–51, 2003.