

## 修飾節に着目した長文における短文化への試み

大野博之<sup>†1</sup> 稲積宏誠<sup>†2</sup>

**概要:** 学生へのアカデミックライティングを指導するうえで、わかりやすい文章となるよう指導する必要がある。学生の書くレポート文でわかりづらい文の要因となるものはいくつか考えられるが、修飾節が連なることで発生する長文がその1つに挙げられる。そこで、この長文を短文へ機械的に変換する試みを行い、文の修正案として提示可能なシステムを目指す。このような短文へ向けた修正の考え方は、日本語の非母語話者を対象とした「やさしい日本語」への取り組みでも、取り上げられていることである。文を短文にするために、文の区切りの目安となる修飾節を検出し、そこを分割点とするのが本システムの基本的な考え方である。そこでICTを活用し、文の修飾節として補足節・連体節・副詞節を自動的に判定し、副詞節を中心とした短文化への候補を提示するシステムを試作した。

**キーワード:** レポート文, 推敲, 修飾節, 副詞節, 長文, 短文

### Attempt to Divide Long Sentences into Short Sentences Focusing on Subordinate Clauses

HIROYUKI OONO<sup>†1</sup> HIROSHIGE INAZUMI<sup>†2</sup>

#### 1. はじめに

学生レポートにおける文章作成能力を育成するために、基礎教養教育の1つとして日本語アカデミックライティングが授業として開かれてきた[1][2]。それに伴い、ライティングセンターを設置する高等教育機関も現れている[3]。こうした中で、指導する教員の負担軽減に活用できるようなICT環境や各種ツールも公開されてきている。

例えば日本語を対象としたリーダビリティに関するツールとしては、次のものが挙げられる。柴崎・原[4]の取り組みでは、9学年もしくは12学年のいずれに該当する文かを予測する。これは、文の平仮名の割合、平均述語数、平均文字数、平均文節数からなる公式によって求められているものである。また、佐藤[5]の取り組みでは、独自に作成した教科書コーパスを基に bigram による言語モデルを作成し、13段階の学年区分を決定している。李[6]の取り組みでは、平均文長、漢語率、和語率、動詞率、助詞率を用いて、重回帰分析による公式を定め、難易度として6段階(初級前半・初級後半・中級前半・中級後半・上級前半・上級後半)を結果としている。これらはいずれも Web サービスとして誰もが利用できる環境で提供されている。

一方、我々は具体的な指導に結び付けることを目標として校正や推敲に対する支援ツールを構築してきた[7]。いわゆる技術文章を対象として、主に文法上の形式的なチェックを行い、修正へ向けた指摘を行うものである。しかし、文のわかりやすさや平易さの支援という点については、文

の構造上の特徴、特に係り受け関係等の特徴を明示することで、自らが書きたい内容と文の構造が合致しているか否かを評価することを前提としてきた。すなわち、「わかりやすい文章」へ修正するための具体的な提案は、いまだ不十分な状況となっている。

他方で、災害時等の情報提供の問題をきっかけとして、日本在住の外国人を対象とした「やさしい日本語」が、盛んに議論されている[8]。ここでの「やさしい」は、非母語話者にとってのわかりやすい文章のことであり、理論的な研究に加えて多くの実践的な取り組みがなされている。こうした取り組みの結果、各自治体でも「やさしい日本語」による情報発信の指針が示されつつある。これらに共通する「やさしい日本語」への指針は、①伝える情報を絞る ②文を短くし簡単な構造にする ③簡単な語彙にする などが挙げられている。

そこで、本研究では、わかりやすい文章へ修正する手段の1つとして「文を短くし簡単な構造にする」という点に着目し、機械的に短文への変換を行い、修正案を提案するシステムの構築を試みた。短文へ変換するためには、元の文を2つ以上に分割する必要がある。そのためには述語を2つ以上伴う複文が処理対象として最適であると考えた。すなわち、複文における修飾節を文の分割点にしようという考えである。ただし、分割後の文同士をつなげる接続関係についても提案を行うためには、修飾節を識別すると同時に、どのような目的で用いられている節なのかも含めて検出する必要がある。

このような考えから、まず修飾節および修飾節の役割を識別したのち、それを利用して文の分割案を提示するシステムを試作することとした。

<sup>†1</sup> 東京医療保健大学  
Tokyo Healthcare University  
<sup>†2</sup> 青山学院大学  
Aoyama Gakuin University

## 2. 修飾節の検出

### 2.1 修飾節の種類

複数の述語からなる複文は、全体の中心となる主節と、その主節を修飾する節（従属節）で構成される。日本語における節のとらえ方は一通りではないため、本システムでは、表 1 のように益岡ら[9]の分類方法をベースにしている。ただし、利用する係り受け解析ツールが既定では並列関係を取得できないため、並列節は除外し、副詞節相当として扱う。すなわち、補足節・連体節・副詞節の 3 種を扱うものとし、これらをここでは修飾節と呼ぶことにする。

表 1 節の分類

| 節種類 | 細分類                                |
|-----|------------------------------------|
| 補足節 | 形式名詞（こと・の・ところ）、疑問表現、引用節            |
| 連体節 | 補足語修飾節、相対名詞修飾節、内容節                 |
| 副詞節 | 時、原因・理由、条件・譲歩、付帯状況・様態、逆接、目的、程度、その他 |
| 並列節 | ※本システムでは並列節は副詞節として扱う               |

補足節は、主節の述語を補う働きをするものである。連体節は、名詞を修飾する節である。副詞節は、述語を修飾したり、文全体を修飾したりする働きをするものである。以下の例の下線部分がそれぞれの節に相当する。

**補足節：**太郎は紅茶が嫌いなことを隠している。

**連体節：**太郎は花子が愛した紅茶をもう見たくない。

**副詞節：**太郎は花子と別れたので、一人で珈琲を飲んだ。

### 2.2 修飾節判定のアルゴリズム

本システムでは、書籍『基礎日本語文法』[9]にしたがって、形態素解析 MeCab[10]と係り受け解析 CaboCha[11]の結果による品詞情報・表記・係り受け関係などから修飾節の判定を行う。また、可能なかぎり細分類に相当する情報を付与する。ただし、1 文節のみで構成される節は存在しないという制約を入れている。図 1 は本システムによって、先に述べた各節の文例を図示した結果である。

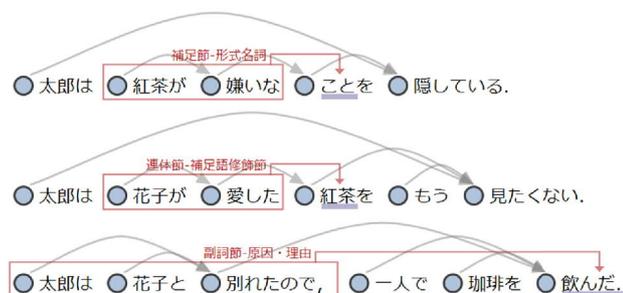


図 1 補足節・連体節・副詞節の図示例

### 2.3 節判定の評価

国立国語研究所のコーパスセンターで公開されている

節境界アノテーション(CBL)[12]の白書(OW)と新聞(PN)データとの比較を次の手順で行った。

- ①節境界アノテーションにおける節ラベルを補足節・連体節・副詞節に分類（表 2）
- ②BCCWJ コーパス[13]から、鍵括弧と丸括弧と伏字用の■記号を含む文および「。」で終了していない文を除外
- ③節境界アノテーションのデータと「②」のコーパス文を結合
- ④本システムで同じ文を処理し、補足節・連体節・副詞節の検出位置を比較

①の分類については、丸山らの分類[14]を参考にした。②の理由は、BCCWJ コーパスにおいて、文中文があった場合の対応である。例えば以下のような例文があると、「紅茶にミルクを入れたい」だけでなく、「太郎は、」と「語った」もそれぞれ文頭フラグが付いており、元の 1 文に復元が難しいためである。

例) 太郎は、「紅茶にミルクを入れたい」と語った。

表 2 節境界アノテーションにおける節ラベルとの対応

| 節種類 | 節ラベル   |
|-----|--|
| 補足節 | 補足節<br>間接疑問節、間接疑問節カドウカ、間接疑問節カドウカ-助詞、間接疑問節-助詞、引用節、引用節ッテ、引用節トノ、引用節ナンテ、引用節-助詞、引用節-補足語、トイウ節  |
| 連体節 | 連体節、連体節カギリノ、連体節ダケノ、連体節タメノ、連体節ナドノ、連体節バアイノ、連体節ホドノ、連体節マデノ、連体節ヨウナ、テカラノ節、テノ節  |
| 副詞節 | 並列節ガ、並列節ケド、並列節ケドモ、並列節ケレド、並列節ケレドモ、並列節シ、並列節シ-助詞、並列節タリ、並列節タリ-助詞、並列節デハ、並列節ヤラ<br>テ節、テカラ節、テカラ節-助詞、テハ節、テモ節<br>条件節カギリ、条件節カギリ-助詞、条件節ケッカ、条件節タラ、条件節ト、条件節トコロデ、条件節ナラ、条件節ナラバ、条件節バ、条件節バアイ、条件節バアイ-助詞、条件節モノ<br>理由節カラ、理由節カラ-助詞、理由節ノデ、タメ節、タメ節-助詞、タメニ節、タメニ節-助詞<br>時間節アト、時間節アトデ、時間節アトニ、時間節アトニ-助詞、時間節アトノ、時間節アト-助詞、時間節イマ、時間節イマ-助詞、時間節イライ、時間節その他、時間節その他-助詞、時間節トキ、時間節トキデ、時間節トキニ、時間節トキニハ、時間節トキノ、時間節トキ-助詞、時間節マエニ、時間節マエノ、時間節マエ-助詞<br>連用節、連用節シニ、ダケ節、ダケ節-助詞、ダケニ節、連用節ッテ、ツツ節、ツツ節-助詞、トカ節、トカ節-助詞、連用節ドコロカ、ナガラ節、ナガラ節-助詞、ナガラモ節、ナド節、ナド節-助詞、ナド節-補足語、譲歩節ニセヨ、連用節ニハ、連用節ニモ、ノニ節、連用節バカリ、連用節バカリ-助詞、ホカ節、ホカ節-助詞、ホド節、ホド節-助詞、ホドニ節、マデ節、マデ節-助詞、マデニ節、マデニハ節、マデハ節、マデモ節、ママ節、ママ節-助詞、ママデ節、ママデ節-助詞、ヨウ節、ヨウ節-助詞、ヨリ節、ヨリ節-助詞、連用節その他、連用中止節、イ形容詞連用節、ズニ節、ズニ節-助詞、ヨウニ節、ヨウニ節-助詞、タメニハ節 |
| 対象外 | 上記以外のラベル（文末* と 対象外）  |

以上の方法で比較した結果、表 3 のとおり約 8 割で、節境界アノテーション(CBL)と同等の結果が得られた。表におけるラベル数とは、節境界アノテーション(CBL)データでのラベル数である。一致数とは、前述のラベルに対して本システムで検出した節のうち一致した数である。

表 3 節境界アノテーション(CBL)と本システムとの比較

| 節種類 | コーパス | 一致度     | 一致数  | ラベル数 |
|-----|------|---------|------|------|
| 補足節 | 白書   | 約 84.6% | 587  | 694  |
|     | 新聞   | 約 78.2% | 1040 | 1330 |
| 連体節 | 白書   | 約 82.7% | 1829 | 2211 |
|     | 新聞   | 約 84.8% | 2512 | 2964 |
| 副詞節 | 白書   | 約 84.2% | 2013 | 2390 |
|     | 新聞   | 約 83.9% | 3450 | 4110 |

なお、一致しなかった原因として、以下のようなものが見受けられた。

- 補足節と連体節のルールが本システムと異なるもの
- 本システムでは、節は 2 文節以上で構成されるもののみを対象とするため、検出できなかったもの
- 文節の区切り方が異なっており、一致しなかったもの

#### 2.4 コーパスデータにおける文の長さや修飾節

以上の修飾節検出方法を用いて、3 つ文章群に対して、5 つの文節数グループごとの平均含有数を算出した。文章群は、BCCWJ の白書コーパスと新聞コーパスおよび、独自に 2008 年～2015 年の間に収集した大学生が書いた文データである。白書・新聞コーパスは、前項で用いたものだけでは文数が少ないため、両コーパス全体を対象としている。3 つ文章群とも同じ条件にするため、鍵括弧・丸括弧を含む文は除外するなど整形し、また 5 文節以下の文も除外してある。その結果、白書は 45579 文、新聞は 12710 文、学生は 17017 文となっている。

表 4 文の長さや修飾節の関係

| 文節         | 種別 | データ割合  | 補足節数 | 連体節数 | 副詞節数 |
|------------|----|--------|------|------|------|
| 6 - 10 文節  | 白書 | 25.95% | 0.13 | 0.37 | 0.64 |
|            | 新聞 | 64.93% | 0.22 | 0.47 | 0.87 |
|            | 学生 | 54.00% | 0.48 | 0.43 | 1.17 |
| 11 - 15 文節 | 白書 | 31.49% | 0.29 | 0.76 | 1.42 |
|            | 新聞 | 26.91% | 0.41 | 0.97 | 1.67 |
|            | 学生 | 30.79% | 0.87 | 0.92 | 2.31 |
| 16 - 20 文節 | 白書 | 22.64% | 0.45 | 1.22 | 2.16 |
|            | 新聞 | 6.53%  | 0.62 | 1.44 | 2.44 |
|            | 学生 | 10.96% | 1.28 | 1.41 | 3.37 |
| 21 - 25 文節 | 白書 | 13.17% | 0.66 | 1.70 | 2.92 |
|            | 新聞 | 1.36%  | 0.88 | 2.02 | 3.16 |
|            | 学生 | 3.26%  | 1.63 | 1.93 | 4.45 |
| 26 - 30 文節 | 白書 | 6.75%  | 0.88 | 2.21 | 3.69 |
|            | 新聞 | 0.27%  | 1.06 | 2.24 | 4.44 |
|            | 学生 | 0.99%  | 2.17 | 2.39 | 5.54 |

結果を表 4 に示す。データの割合から、いずれの文章群においても、11～15 文節の文が多くを占めていることがわかるが、白書は 16 文節以上も多い一方、新聞はほぼ 15 文節以下に抑えられていることがわかる。学生も、文の長さでいえば、新聞と同じように 15 文節以下が大部分を占めている。しかし、いずれの長さにおいても、補足節と副詞節に関して、白書や新聞より多用しがちであることが認められる。例えば学生では 11～15 文節でも平均 2.31 個の副詞節で文を構成している。そこで、本システムでは、特に顕著である副詞節を中心に短文化を試みる。

### 3. 文の短文化

本システムにおける文の短文化の目的は、長文を分割することである。したがって、短文化したあとの文の長さが短くなってしまふ場合は、基本的に対象外とする。すなわち、図 2 の例 1 のように分割後もある程度の長さであれば良いが、例 2 のように分割後の片方が極端に短くなるような場合は、対象外とするということである。ただし、分割後の長さがどの程度が適当であるかは、文のタイプや読み手によって変わることもあるため、本システムではオプションとしている。

例 1)  
**太郎が飲んだその日の珈琲は苦かったが、少し落ちつくことができた。**  
 太郎が飲んだその日の珈琲は苦かった。  
 少し落ちつくことができた。

例 2)  
**黒いだけが、今の太郎にはそんな色の珈琲が合っていた。**  
 黒いだけだ。  
 今の太郎にはそんな色の珈琲が合っていた。

図 2 文の短文化例

分割の方針として、3 種のタイプに対して行う。まず、接続助詞による分割、次に一部の接続表現による分割、そして副詞節による分割である。以下、これらの分割について説明する。

いずれの場合も文の分割点を検出し、分割後の接続関係を付与し、2 つの文に整形するという方法が基本的な流れとなる。接続関係は「順接、逆接、添加、並立、説明、対比、転換、無し」の 8 種とした。

また、2 文に分割する際に、考え方として 2 つの方法がある。1 つは分割点で単純に分割する方法であり、もう 1 つは分割点で分割するものの係り受け関係を維持する場合である。つまり図 3 のように 2 通りの分割ができるということである。これは係り受け解析の結果に影響するものであるため、本システムでは、両方の結果を出力するものとした。

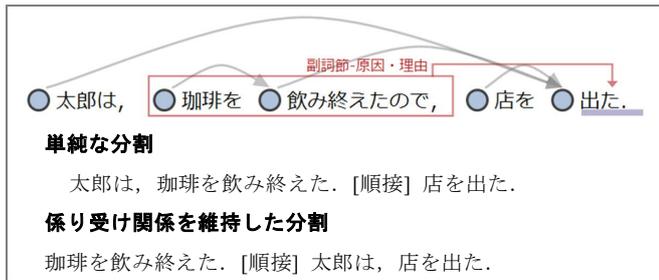


図 3 分割時の文の切り方の違い

### 3.1 接続助詞による分割

接続助詞は、用言（用言に準ずるもの含む）に付いて、その後続く文との意味上の関係を示す助詞である。副詞節の一部として該当するが、ここでは別個扱う。手順は次のとおりであり、対応する接続助詞と接続関係は表 5 のとおりである。ここでの接続助詞は IPA 辞書に登録されているものを前提としている。なお、「と」と「つつ」に関しては、これを含む文を手で見て分割しづらかったため除外してある。

- ①文中で接続助詞を検出
- ②接続助詞を含む文節がすぐ後ろに係る場合は対象外
- ③接続助詞の種類から、接続関係を決定

表 5 接続助詞と接続関係の対応

| 接続助詞                                      | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|---|----|----|----|----|----|----|----|----|
| および                                       |    |    | ○  |    |    |    |    |    |
| が、けれど、けれども、けど、けども、て、で、                    | ○  | ○  |    | ○  |    |    |    |    |
| から、ので                                     | ○  |    |    |    |    |    |    |    |
| からには、さかい、じゃ、だって、ちゃ、ちゃあ、とも、ども、なり、や、やいなや、んで |    |    |    |    |    |    |    | ○  |
| し   |    |    |    | ○  |    |    |    |    |
| ど、どころか、のに、ものの                             |    | ○  |    |    |    |    |    |    |
| ながら                                       |    | ○  | ○  | ○  |    |    |    |    |
| ば   | ○  |    |    | ○  |    |    |    |    |

### 3.2 一部の接続的表現による分割

白書で顕著である用言と用言を結ぶ表現について分割を行う。具体的には「用言+とともに」、「用言+ほか」、「用言+ことより」、「用言+ことから」の 4 種である。

BCCWJ の白書 (OW) コーパスに対して、鍵括弧や丸括弧を含む文を除くなど整形して得た 51239 文のうち、80 文字以上のものは 18702 文ある。ここに上記の表現が含まれるのは、順番にそれぞれ 2581 文、854 文、635 文、730 文ある。一方、80 文字未満である残り 32537 文では、順番にそれぞれ 648 文、251 文、225 文、247 文となっており、上記の表現は長文ほど用いられやすい。

なお、比較的短文が多い新聞 (PN) コーパスでは、同様に整形して得た 18163 文すべてを対象にしても、上記の表現は順番にそれぞれ 21 文、41 文、1 文、49 文というように出現数は少ない。

### 3.3 副詞節による分割

書籍『基礎日本語文法』によると、「副詞節は、述語の修飾をしたり、文全体を修飾したりする働きを持つ」とされている。したがって、補足節や連体節とは違い、副詞節自体を取り出して大きく変更することなく文に整形しやすい。また、副詞節の表す意味が接続関係と直結しやすいことから、分割処理の対象とした。一部、接続助詞による分割と重複する。副詞節が表す意味は、時、原因・理由、条件・譲歩、付帯状況・様態、逆接、目的、程度などがある。

#### 時を表す副詞節 (表 6)

- ①「場合」「とき」で分割し、前半は「基本形+とする」を追加。後半は「その」を追加。
- ②「まで」で分割し、前半は「ある or あった」を追加。
- ③「後」「際」で分割、後半は「その」を追加。
- ④「最中」「あいだ」で分割し、前半は「基本形」を追加。後半は「その」を追加。

表 6 副詞節分割時の接続関係 (時)

| 表現       | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|----------|----|----|----|----|----|----|----|----|
| 場合、とき    |    |    |    |    |    |    |    | ○  |
| まで       |    |    | ○  |    |    |    |    |    |
| 後、際、最中、間 |    |    |    |    |    |    |    | ○  |

#### 原因・理由を表す副詞節 (表 7)

- ①「ため」「から」で分割し、前半は「である」を追加。前半と後半を逆に配置。
- ②「結果」で分割。後半は「その」を追加。
- ③「なく」で分割し、「ない」に修正。
- ④「から」で分割し、前半は「である」を追加。前半と後半を逆に配置。
- ⑤「だけに」で分割。
- ⑥「もので」で分割。「もので」を除去するものと「ものである」の 2 パターンを作成。

表 7 副詞節分割時の接続関係 (原因・理由)

| 表現    | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|-------|----|----|----|----|----|----|----|----|
| ため、から |    |    |    |    | ○  |    |    |    |
| 結果    |    |    |    |    |    |    |    | ○  |
| なく    |    |    |    |    |    |    |    | ○  |
| から    |    |    |    |    | ○  |    |    |    |
| だけに   | ○  |    |    |    |    |    |    |    |
| もので   | ○  |    |    |    |    |    |    |    |

**条件・譲歩を表す副詞節 (表 8)**

- ①「ば」で分割し、前半は「基本形+とする」を追加。

表 8 副詞節分割時の接続関係 (条件・譲歩)

| 表現 | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|----|----|----|----|----|----|----|----|----|
| ば  | ○  |    |    |    |    |    |    |    |

**付帯状況・様態を表す副詞節 (表 9)**

- ①「代わり」で分割し、「基本形」に修正。後半は「その」を追加。  
②「ついでに」で分断。

表 9 副詞節分割時の接続関係 (付帯状況・様態)

| 表現   | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|------|----|----|----|----|----|----|----|----|
| 代わり  |    |    |    |    |    |    |    | ○  |
| ついでに |    |    |    |    |    |    |    | ○  |

**逆接を表す副詞節 (表 10)**

- ①「のに」「にもかかわらず」で分断。ただし「名詞+の」の場合は「である」を追加。

表 10 副詞節分割時の接続関係 (逆説)

| 表現         | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|------------|----|----|----|----|----|----|----|----|
| のに、にもかかわらず |    | ○  |    |    |    |    |    |    |

**目的を表す副詞節 (表 11)**

- ①「べく」で分割し、前半は「ためである」を追加。前半と後半を逆に配置。

表 11 副詞節分割時の接続関係 (目的)

| 表現 | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|----|----|----|----|----|----|----|----|----|
| べく |    |    |    |    | ○  |    |    |    |

**程度を表す副詞節 (表 12)**

- ①「なく」で分割し、「ない」に修正。  
②「ぐらい」「くらい」で分割し、前半は「ある」を追加。

表 12 副詞節分割時の接続関係 (程度)

| 表現      | 順接 | 逆接 | 添加 | 並立 | 説明 | 対比 | 転換 | 無し |
|---------|----|----|----|----|----|----|----|----|
| なく      |    |    |    |    |    |    |    | ○  |
| ぐらい、くらい |    |    | ○  |    |    |    |    |    |

**4. 評価**

**4.1 白書および学生データの短文化**

2.4 で使用した白書コーパスおよび学生データに対して、文を分割した結果、表 13 のとおりとなった。16 文節以上については概ね 5 割以上の文に対して、短文化への提案が実現できている。なお、今回の分割の条件は、「分割後の文が 6 文節以上」としている。そのため、6~10 文節の文については、結果から除外した。

表 13 本システムで短文化の提案できた文数

| 文節         | 種別 | データ数    | 副詞節を含む文 | 分割できた文 |
|------------|----|---------|---------|--------|
| 11 - 15 文節 | 白書 | 14352 文 | 11940 文 | 2290 文 |
|            | 学生 | 5239 文  | 4917 文  | 906 文  |
| 16 - 20 文節 | 白書 | 10320 文 | 9585 文  | 4363 文 |
|            | 学生 | 1865 文  | 1824 文  | 854 文  |
| 21 - 25 文節 | 白書 | 6001 文  | 5825 文  | 3625 文 |
|            | 学生 | 554 文   | 547 文   | 330 文  |
| 26 - 30 文節 | 白書 | 3077 文  | 3025 文  | 2137 文 |
|            | 学生 | 169 文   | 169 文   | 119 文  |

**4.2 機械翻訳による短文化への影響**

近年では、長文に対する機械翻訳の精度も向上してきている[15]が、従来では苦手なものとされてきた。そこで、文の分割前の原文と分割後の文を機械翻訳にて英訳し、それを再度和訳することによって、文脈的な違いが起きないかどうか、いくつかのサンプルで実施した。機械翻訳には、みらい翻訳[16]のお試し翻訳を使用させていただいた。以下に 3 つのパターンの例を示す。

**(1) 分割してもしなくても特に問題がない例**

**原文** 高齢者が働ける環境が整えば、高齢者は働けて、お金を趣味や娯楽に使う余裕が出てくる。

**英訳** If an environment is created in which elderly people can work, they will be able to work and spend their money on hobbies and entertainment.

**再和訳** 高齢者が働ける環境が整えば、仕事もできるし、趣味や娯楽にお金を使うこともできる。

**分割** 高齢者が働ける環境が整えば、高齢者は働ける。お金を趣味や娯楽に使う余裕が出てくる。

**英訳** If an environment is created in which elderly people can work, elderly people can work. They can afford to spend money on hobbies and entertainment.

**再和訳** 高齢者が働ける環境が整えば、高齢者も働くことができる。彼らは趣味や娯楽にお金を使う余裕がある。

**(2) 分割することで翻訳が改善する例**

**原文** お店の食材などを悪ふざけで使って写真を撮りその

様子を SNS にアップロードする行為は常識があったら普通やらないと思うが彼らは面白半分だったり、かっこつけた自慢だったり、人とは違って目立ちたいんだといった感情からバイトテロを引き起こすと考えた。

**英訳** They thought that if they had common sense, they wouldn't usually take photos of the store's food and upload them to social networking sites, but they thought that it was just for fun, or because they wanted to stand out and stand out.

**再和訳** 常識があれば、普段はお店の料理の写真を撮って SNS にアップすることはないだろうと思っていましたが、それは単なる遊びだとか、目立ちたいからだと思っていました。

**分割** お店の食材などを悪ふざけで使って写真を撮りその様子を SNS にアップロードする行為は常識があったら普通やらないと思う。彼らは面白半分だったり、かっこつけた自慢だったり、人とは違って目立ちたいんだといった感情からバイトテロを引き起こすと考えた。

**英訳** If you have common sense, you don't usually take a picture of a store's food as a prank and upload it to SNS. They thought it was because they wanted to be funny, cool bragging, or stand out.

**再和訳** 常識があれば、お店の食べ物をいたずらで撮影して SNS にアップすることはあまりありません。彼らはそれを、面白がったり、クールな自慢をしたり、目立つことをしたいからだと思っていました。

### (3) 分割することで翻訳時に意味が変化する例 (下線部)

**原文** 高学年にもなれば、塾に行きだす子どももいるので、帰宅の際の迎えを親にメールや電話で頼むことができる。

**英訳** In the upper grades, some children start going to cram schools, so they can ask their parents to pick them up when they return home by e-mail or phone.

**分割** 高学年にもなれば、塾に行きだす子どももいる。帰宅の際の迎えを親にメールや電話で頼むことができる。

**英訳** Some children start going to cram schools when they reach the upper grades. You can ask your parents to pick you up by email or phone when you get home.

## 5. おわりに

学生が記述する文は、特に補足節と副詞節による長文が散見されるため、本システムでは、修飾節を文の分割点として、長文を短文化する試みを行った。主に副詞節を中心とした分割であったが、16 文節以上の文に対しては概ね 5 割以上に対して、短文化を提案する環境を構築できた。

その中で、短文化によって文の構造がシンプルになり、

わかりやすくなることもあるが、一方で分割後の情報の欠如によって、意味が不明確になる点や分割箇所によっては、文が不自然となるケースも見受けられた。また、分割の根拠となる節が複数あった場合に、どのような分割の仕方が最適なのかという問題も残っている。

今後の課題としては、どのような場合にわかりやすくなるのかの指標を求めること、分割した際に不足する情報の補い方のルールを明確にすること、学生が多用する補足節を含む文の短文化への取り組みが挙げられる。また、今回の評価では、分割後 5 文節以下になるものは分割しないという設定でおこなっているが、どの程度の長さで分割するかによって得られる結果は異なる。したがって、どのような長さが最適かも合わせて検討していく必要がある。

## 参考文献

- [1] 中島利勝, 塚本真也. 知的な科学・技術文書の書き方. コロナ社, 1996
- [2] 塚本真也. 知的な科学・技術文書の徹底演習. コロナ社, 2007.
- [3] 小林 至道, 中竹 真依子. ライティングセンター運営上の現状と課題: 青山 AWC の立ち上げから 1 年間の量的データにもとづく分析. 青山インフォメーション・サイエンス 青山学院大学附置情報メディアセンター 46 巻 1 号, pp4-9, 2013
- [4] 柴崎秀子, 原信一郎. 12 学年を難易尺度とする日本語リーダビリティ判定式. 計量国語学, 27-6, pp.215-232, 2010.
- [5] 佐藤理史. 均衡コーパスを規範とするテキスト難易度測定. 情報処理学会論文誌, Vol.52, No.4, pp.1777-1789, 2011.
- [6] 李在鎬. 日本語教育のための文章難易度研究. 早稲田日本語教育学, Vol.21, pp.1-16, 2016.
- [7] 又平恵美子, 竹内純人, 大野博之, 稲積宏誠. 文章作成支援ツールによる日本語文章力育成. 私立大学情報教育協会 ICT 活用教育方法研究 第 13 巻 第 1 号 p.16-20, 2010.
- [8] 佐藤 和之. 在住外国人 300 万人・訪日外国人 4000 万人時代の安全を支える「やさしい日本語」: 総務省消防庁『避難誘導のあり方』ガイドラインでの「やさしい日本語」採用の理由と言語学的根拠. 消防防災科学センター 消防防災の科学 (137), 49-56, 2019.
- [9] 益岡隆志, 田窪行則. 基礎日本語文法・改訂版. くろしお出版, 1992.
- [10] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto. Applying Conditional Random Fields to Japanese Morphological Analysis. Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.
- [11] 工藤拓, 松本裕治. チャンキングの段階適用による係り受け解析. 情報処理, Vol 43, No. 6, pp.1834-1842, 2002.
- [12] 佐藤理史, 丸山岳彦. 節境界認定に関する諸問題 国立国語研究所 第 8 回 コーパス日本語学ワークショップ 予稿集, 225-232, 2015.
- [13] 現代日本語書き言葉均衡コーパス(BCCWJ). [http://www.ninjal.ac.jp/corpus\\_center/bccwj/](http://www.ninjal.ac.jp/corpus_center/bccwj/), (参照日 2020.2.20)
- [14] 丸山岳彦・佐藤理史・夏目和子. 現代日本語における節の分類体系について. 言語処理学会第 22 回年次大会発表論文集, pp.1113-1116, 2016.
- [15] 国立研究開発法人情報通信研究機構 (NICT). ニューラル機械翻訳で音声翻訳アプリ VoiceTra が更なる高精度化を実現. <https://www.nict.go.jp/press/2017/06/28-1.html>, (参照日 2020.2.20).
- [16] 株式会社みらい翻訳. TOEIC900 点以上の英作文能力を持つ深層学習による機械翻訳エンジンをリリース. <https://miraitranslate.com/news/313>, (参照日 2020.2.20).