

# 翻訳のための起点文書分析：文献レビューの枠組み

宮田 玲<sup>†</sup> 宮内 拓也<sup>‡</sup> 影浦 峯<sup>‡</sup>

<sup>†</sup>名古屋大学 <sup>‡</sup>東京大学

## 1 はじめに

“Connect the engine wire to the cylinder head cover sub-assembly with the bolt.”<sup>1</sup>

この文だけを見せられて、「日本語に翻訳してください」と依頼されたら、翻訳者であればまずは「これだけでは翻訳できません」と答えるだろう。ここで、「前後の文も3文ずつお見せします」と言われても、翻訳できない状況には変わりない。もちろん、このように1文（ないし数文）だけ渡されて翻訳するという案件は現実の産業翻訳ではありえないだろうが、機械翻訳のユースケースを考えるとまさにこのような状況で「翻訳」が行われている。例えば、最新のニューラル機械翻訳を使えば、「エンジンワイヤーをシリンダーヘッドカバーサブアセンブリにボルトで接続します。」<sup>2</sup>と英文和訳の試験であればそれなりに点が取れる結果を出せるが、翻訳としての妥当性は担保されていない。

仮に、ここで妥当な翻訳は「ボルトでエンジンワイヤーをシリンダーヘッドカバー SUB-ASSY に接続する。」であるとすると、少なくとも言語仕様としては、以下が決まっていなければならない。

- “connect”, “engine wire”, “cylinder head cover sub-assembly”, “bolt” の日本語訳はそれぞれ「接続する」「エンジンワイヤ」「シリンダーヘッドカバー SUB-ASSY」「ボルト」である
- 最初に補足要素（使用部品）、次にヲ格要素、最後にニ格要素を並べる
- 使用部品は「～で」の形で使う
- 文末は常体の平叙文「～する」の形を使う

そしてこのような言語仕様を決めるためには、例えば、以下のような判断がなされる必要がある<sup>3</sup>。

- 工業分野の技術文書では、“connect” の日本語訳は「接続する」が標準的である
- A 社の対訳用語集を見ると、“engine wire”, “cylinder head cover sub-assembly” の日本語版

はそれぞれ「エンジンワイヤ」「シリンダーヘッドカバー SUB-ASSY」である

- A 社の X という自動車の修理書に書かれている文を見ると、ヲ格要素はニ格要素に先行しており、使用部品は文の最初に書かれる傾向があるので、最初に使用部品、次にヲ格要素、最後にニ格要素を並べるのが標準的である
- A 社のスタイルガイドによると、修理書内の一連の操作手順は、常体の平叙文「～する」の形を使う必要がある

さらに、このような判断をするためには、原文は「A 社の B という自動車の修理書の C という修理項目中の修理手順に記載された1つの操作説明」であることを、知っていなければならない。

恣意的かつ限定的な例であったが、以上の説明を逆から辿ることで、妥当な翻訳を担保するプロセスの一端が明らかになってくる。一般化してしまえば、「翻訳対象が何であるかを把握し、それに応じて適切な知識を参照し、テキストに対して適切な操作を加え、所定の要件を満たした翻訳結果を作成する」ということになる。しかし、ここで最大の問題は、この翻訳プロセスを具体的に説明・共有する上で十分な解像度を備えた概念・言葉が整理されていないことである。

このような背景のもと、我々は「翻訳とは何か？」という大元の問いを「翻訳とは、何に対し、どのような要素・属性を同定し、それにどのような操作を加え、どのような基準でその操作を終了させる行為か？」と読み替え、翻訳プロセスで何が行われているかを具体的に言語化し共有することを目的とした研究プロジェクトを進めている [1]。現在は、この問いの前半部の「何に対し、どのような要素・属性を同定するのか」に答える手がかりとして、翻訳学やその関連分野において、起点言語の文書のどのような要素・属性が言及されてきたかを、文献レビューにより洗い出す作業を進めている。

この作業の成果は、言語処理分野にも示唆を与える。特に近年、機械翻訳研究においては、1文を超える単位の翻訳の研究が増えている。例えば、2013年から

<sup>1</sup>PRIUS 修理書 (トヨタ自動車株式会社, 2017年11月北米版) から引用。下で妥当な翻訳例として示した文は、PRIUS 修理書の日本版から引用。なお、実際には北米版から日本版が作られたわけではなく、あくまで例として取り上げた。

<sup>2</sup>Google 翻訳結果 (2020年1月14日取得)

<sup>3</sup>あくまで例であり、これが実際に正しいわけではない。

開催されている Workshop on Discourse in Machine Translation (DiscoMT) では、以下のような文を超えたテキストの属性が例として挙げられている [2, p.iii]。

- 文書全体に関する属性（スタイル、レジスター、読解レベル、ジャンル）
- トピック的・機能的な部分構造
- 談話の一貫性（文、節、参照形式で明示的・暗黙的に表現される）
- 照応表現や省略表現（話者が先行する談話文脈を利用し、後続の情報を簡潔に示す）

我々は、これらも包摂するような文書の要素・属性の全体像を具体的かつ網羅的に明らかにすることを目指しており、その知見の一部は機械翻訳研究にも資するだろう。本稿では以下、いくつかの概念を整理した上で、文献レビューの枠組みを、予備的な調査結果とともに報告する。

## 2 概念整理

### 2.1 文書

前節で例示したように、翻訳者が翻訳するのは、文ではなく、**文書 (document)** である。これは、翻訳作業上の単位の話ではない。たとえ、文書中の 1 文のみが翻訳対象だとしても、また翻訳作業に際して文書の全体を実際に読まないにしても、その特定の文書の存在を前提として、翻訳はなされる。

まず本稿では、暫定的な定義として、文書とは「社会において特定のコミュニケーション上の機能・役割を持つ内容的に完結したテキスト」とする<sup>4</sup>。例えば、自動車の修理書に「エンジンオイルの交換」という見出しの文書があったとする。この文書の第一義的なコミュニケーション上の機能・役割（あるいは目的）は、エンジンオイルを交換しようとしている読み手に、その方法を間違いなく伝え、作業の完遂を促すことである。そして、その機能を果たすためには、その文書は内容的に完結している必要がある<sup>5</sup>。

この定義に照らすと、文書は階層的に存在していてもよい。例えば、10 ページの電化製品マニュアルがあったとき、「その製品の使用方法を説明する」機能を果たしている全体を 1 文書と考えることもできるし、その中の「電源が入らないという課題を解決する方法

を説明する」機能を果たしている部分を 1 文書と考えることもできる。ただし、マニュアルから無作為に文を取り出した場合、それは文書を構成しえない。別の言い方をすれば、無作為抽出文で、テキストコーパスは作れるかもしれないが、文書コーパスは作れない。

ここで、翻訳対象である起点言語における文書を**起点文書 (source document: SD)**、翻訳結果である目標言語における文書を**目標文書 (target document: TD)** と呼ぶことにする。翻訳学においては、起点テキスト (source text: ST) と目標テキスト (target text: TT) の用語が一般的であるが、我々は「文書」であることを明示する場合は積極的に SD, TD を用いる。

### 2.2 文書プロフィール

それでは、翻訳において、SD はどのように扱われるのであろうか。SD を文の集合とみなし、さらに文を単語の集合とみなすような、言語学的な観点による分析だけでは、SD の社会的な属性の多くが捨象されてしまうことは言うまでもない。SD を、社会において固有の機能・役割を持った代替不可能なものとみなしながらも、それを操作可能なものとして翻訳するためには、SD (や SD 内の要素) の属性を一定の観点から分析し、その属性を適切に調整しながら、翻訳先の社会に移し替え、TD として実現する必要がある。翻訳学の用語法を参考にしつつ、SD 分析の結果を **SD プロファイル**、それを目標言語に移し替えたものを **TD プロファイル**、両者を総称して**文書プロフィール**と呼ぶこととする。

ここで文書の属性とは、「その文書がどのようなものであるかを示したもの」と広く捉えておく。属性の基本構造は、**属性名と属性値**の組み合わせであり、本稿では、{ 属性名 : 属性値 } と表記する。例えば、情報科学分野の学術論文であれば { 分野 : 情報科学 }, { タイプ : 学術論文 } と表すことができる。

以上を踏まえると、文書プロフィールとは、当該文書に関する複数の属性を分類して列挙したもの、と捉えることができる。なお、翻訳プロセスにおいて、SD プロファイルと TD プロファイルは、同一の属性群を持つわけではない。例えば、スウィフトの *Gulliver's Travels* を日本語の小学生向け絵本に翻訳する場合を考えると、{ 言語 : 英語 }, { タイプ : 小説 }, { 対象読者 : 一般 } という SD 属性は、{ 言語 : 日本語 }, { タイプ : 絵本 }, { 対象読者 : 小学生 } という TD 属性に変わりうる。

<sup>4</sup>この定義は [3, p.27] で言及されている “documents as the units of text that have a function in communication” という捉え方に概ね近い。歴史・社会的な文書の固有性に関する議論は、[1] を参照されたい。

<sup>5</sup>実際に完結しているかは問題ではない。人は文書に接するとき、それが「有限の範囲で完結しているべきもの」とみなすからこそ、そもそも「完結しているかどうか」を判断できる。

スコポス理論では、翻訳目的（スコポス）が第一義的に重要とされ、続いてTDに求められる機能が決まり、それに応じてSDが参照される[4]。これに従えばまずTDプロフィールが翻訳の意思決定に重要であり、それと対応した形でSDプロフィールが参照されることとなる。しかし、注意すべきは、TDプロフィールは、TDの作成に先行して、あくまで要件的に定義されるということである。そして、TDプロフィールを定義する上では、常に「何を」対象とするかが前提としてされていなければならない、TDが具体物として実体化されていない段階においては、具体物のあるSDとそこから作成したSDプロフィールに決定的に依存する。いずれにせよ、SDがどのようなものであるかの分析的把握は翻訳プロセスの出発点となる重要なものである。

### 3 SD分析に向けた文献レビューの枠組み

ここまでの議論を踏まえると、「何に対し、どのような要素・属性を同定するのか」という問いで規定される翻訳プロセスは、「SDに対して、SD分析を行うことで、SDプロフィールを作成する」と読み替えることができる。このプロセスを翻訳者・翻訳学習者が明示的に実行できるように、さらには、ある程度自動化できるようにするには、(1)SDプロフィールの全体構造の定義と(2)SD分析方法の定式化、が必要である。我々は出発点として(1)を対象に、SDプロフィールにはどのような属性が記述されるべきか、という問いのもと、翻訳に関わるSD属性の網羅的列举と分類を行うこととした。

SD属性を洗い出すために、本研究では文献レビューの方法をとる。1節で示したような具体例をもとに、そこで認定されるSD属性を抽出し分類する方法も考えられるが、事例や作業者に偏る可能性が高く、作業コストもかかるため、比較的知見のまとまった文献からの抽出を試みる。翻訳学においては、体系的であるか否かに関わらず、様々なSD分析方法が言及されてきた。SD分析に特化した教科書も存在する[5]。しかし文献間で、認定する属性の観点や粒度が揃っておらず、必ずしも共通の語彙が定まっているわけではない。また文書構造や言語的特徴に関する言及は必ずしも十分ではなく、見落とされている属性もあると考えられる。そこでSD分析に向けた第一歩としては、翻訳学を中心にその関連分野にも対象を広げながら、文献に言及された文書属性を収集・分類していくことが有用である。

### 3.1 対象文献

大きく、翻訳学、テクニカルライティングなどの執筆技術論、言語学、図書館情報学の分野の文献を対象とする。翻訳学は、本レビューの直接の対象領域であり、翻訳のためのSD属性の抽出が期待できる。執筆技術論は、「文書」「テキスト」を対象に、その構造・メカニズムの解明、さらに執筆に特化した言語技術を扱ってきた。主にSDの新規作成を扱う点で、SDが所与のものとしてされる翻訳と異なるが、文書を捉える視点は基本的に共通している。言語学は、「言語」を対象に、その機能の記述・解明を主な対象としている。分析のための語彙も比較的確立しており、翻訳における言語要素の分析に援用できると考えられる。またコーパス言語学では、ジャンルやレジスターといった「文書」や「テキスト」に関する特徴も扱われている。図書館情報学は、主に「図書」「文書」を対象に、その歴史的・社会的位置づけを扱ってきた。例えば、図書館の図書に付与される書誌情報（書名、著者、出版年、主題分類、件名など）は、SD属性とみなすことができるように、関連する知見が得られると期待できる。

翻訳学の文献には、教科書・概説書、事典・ハンドブック、専門書・学術論文、実務家向けの翻訳指南書、翻訳者によるエッセーなど様々なジャンルがある。なるべく網羅的な属性体系の構築を行うために、最終的にはこれらを広く対象とするが、まずは全体像の大きな把握のために、教科書・概説書を対象とする。

### 3.2 文献レビューの手順

以下の手順で、文書属性を抽出・整理し、出典とともにまとめる。

1. 文書属性に言及している箇所を抽出し、属性名と属性値を同定する
  - (a) 属性名は名詞句で記述する
  - (b) 属性値は記載例をいくつか抽出する
2. 属性の定義が書かれていれば抽出する／明示的に書かれていなければ内容を要約し記録する
3. 属性を類型化する
4. 属性間の依存関係を整理する

### 3.3 予備調査

属性の多様性を大まかに把握するため、予備調査として、翻訳学における様々な理論を概説した入門書[4, pp.110–137]を対象に文書属性を抽出した。類似のものを統合し表現を一部修正した結果を、大きく文書の外的な属性と内的な属性で分けて、以下に示す。

## 外的属性

- ├─ タイプ (Reiss 1971)
  - │ : 情報型/表現型/効力型
- ├─ ジャンル・種類 (Chesterman 1989)
  - │ : 詩/伝記/操作説明書/広告など
- ├─ 媒体 (Nord 1997)
- ├─ 主題分野 (Reiss 1971)
- ├─ 意図されたテキスト機能 (Nord 1997)
- ├─ 言語機能 (Reiss 1971)
  - │ : 叙述/表出/訴え
- ├─ 言語特性 (Reiss 1971)
  - │ : 論理的/審美的/対話的
- ├─ テキスト焦点 (Reiss 1971)
  - │ : 内容重視/形式重視/訴求重視
- ├─ 情動的含意 (Reiss 1971)
- ├─ 状況 (Reiss 1971)
- ├─ 時 (Reiss 1971; Nord 1997)
- ├─ 場所 (Reiss 1971; Nord 1997)
- ├─ 送り手 (Holz-Mänttari 1984 など)
  - │ ─ 発起者
  - │ ─ 依頼者
  - │ ─ 作成者
- ├─ 受け手 (Reiss 1971; Nord 1997)
- ├─ 動機 (Nord 1997)
  - │ ─ 執筆された理由
  - │ ─ 翻訳される理由

## 内的属性

- ├─ 題材 (Nord 2005)
- ├─ 意味・内容 (Reiss 1971; Nord 2005)
  - │ ─ 内包的意味
  - │ ─ 結束性
- ├─ 前提 (Nord 2005)
- ├─ テキスト構成 (Nord 2005)
  - │ ─ マクロ構造
  - │ ─ ミクロ構造
- ├─ 語彙 (Reiss 1971; Nord 2005)
  - │ ─ 方言
  - │ ─ レジスター
  - │ ─ 専門用語
- ├─ 文法 (Reiss 1971)
- ├─ 文体 (Reiss 1971)
- ├─ 超分節的特徴 (Nord 2005)
  - │ ─ 強勢
  - │ ─ リズム
  - │ ─ 文体上の理由による休止
- ├─ 非言語要素 (Nord 2005)
  - │ ─ 書体
  - │ ─ イラスト

現時点では網羅的とは言えず、粒度も粗いが、幅広い属性が抽出できた。これをベースに、関連分野にも調査を広げながら、分類を拡大・精緻化していく。また例えば、意図されたテキスト機能と言語機能の違いや、外的属性の主題分野と内的属性の題材の違いなど、概念的に整理が不十分な箇所の検討も進める。

なお今後、SD 分析方法を定式化する上では、属性値の範囲と認定方法に関する指針が必要となる。例示された属性値を収集するだけでなく、標準的・規範的な枠組みを調査しながら、取り得る値やその書き方を明確化したい。

## 4 おわりに

繰り返しになるが、我々は「翻訳とは何か？」という問いを、「翻訳とは、何に対し、どのような要素・属性を同定し、それにどのような操作を加え、どのような基準でその操作を終了させる行為か？」として、妥当な翻訳を保証する翻訳プロセスを、具体的な言葉で記述することを目指している。本稿では、とりわけ「何に対し、どのような要素・属性を同定し」の部分で、「起点文書 (SD) に対し、SD や SD 内の要素を分析し、一定の SD 属性を同定する」と捉え直し、関連概念を整理した上で、文書属性の体系的把握のための文献レビューの方法について述べた。

大元の問題意識には、翻訳学・産業翻訳・言語処理の各分野の研究者・翻訳実務者・翻訳提供者・翻訳利用者が共通で使える、翻訳について語るための言葉があまり整備されていないということがある。文献レビューは現在進行形であり、議論も中途半端ではあるが、問いと概念を提起しておくこと自体が重要と考え、現在の我々の考えを文書化した。ここで提起した用語法や文献レビューの枠組みは叩き台として、今後議論し修正していく予定である。

謝辞 本研究の一部は科研費 (19H05660, 19K20628, 19K23073) の支援を受けた。

## 参考文献

- [1] 影浦峽「人間の翻訳と機械の翻訳 (1): 翻訳者は何を翻訳しているか?」AAMT ジャーナル, 71: 14-19, 2019.
- [2] *Proc. of the Workshop on Discourse in Machine Translation*, 2017.
- [3] Sager, J. C. "Text Types and Translation," *Text Typology and Translation*. John Benjamins, 25-41, 1997.
- [4] マンデイ, J. 『翻訳学入門』鳥飼玖美子 訳, みすず書房, 2009.
- [5] Nord, C. *Text Analysis in Translation: Theory Methodology, and Didactic Application of a Model for Translation-Oriented Text Analysis*. Brill Rodopi, 2006.