

# 翻訳における QA 記述の分析: 言語学的カテゴリーを手掛かりに

宮内 拓也<sup>†</sup> 影浦 峯<sup>‡</sup>

<sup>†</sup> 東京大学 大学院教育学研究科

<sup>‡</sup> 東京大学 大学院情報学環

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

## 1 はじめに

翻訳への需要が増加の一途を辿る中で、実務翻訳における品質管理基準も整備されてきている。しかし、現状では「翻訳」が具体的なプロセスとして、十分に高解像度で記述されているとは言えない。これに起因して、「翻訳」をめぐる関連するアクター間の理解に齟齬をきたしているという問題が表出している。翻訳プロセスを外在的なものとして高解像度で語ることに必須である。また、翻訳の品質管理は翻訳の出口側の最重要課題であると位置づけられるため、品質管理について高解像度で語ることが求められる。

そこで、本稿では、翻訳プロセスにおける品質保証/品質評価(以下, QA)のスキームを対象に、その記述解像度の分析を行う。QA スキームを取り上げるのは、それが翻訳プロセスや翻訳成果物に関して語るべき重要なポイントを体系的に示しているとみなせるからである。各 QA の記述は、翻訳プロセスをモデル化する際には、より具体的な操作として記述し直される必要があるが、その第一歩として、本稿では QA について扱うスキーム内での記述と言語学的なカテゴリーとの対応を記述することで、QA の記述解像度を分析する。

翻訳とは言語学的なプロセスに収まるものではないが [1, 2], 翻訳とは文書を対象に行われる行為であり、文書は(主に)言語で記述される。(理論)言語学は、射程とする範囲で言語についての体系を構築しており、言語、およびその下位要素について語るためのカテゴリーを提供している。言語学的なカテゴリーは概念が固定的であり、比較的定義が容易であることから、本稿では言語学的なカテゴリーを軸に分析を行う。

以下、2 では QA について、3 では翻訳品質の評価の際の観点について、それぞれ概要を述べる。4 では、各 QA の記述がどのような言語学的カテゴリーを対象としているのかを記述する。5 は本稿全体のまとめである。

## 2 QA スキームについて

翻訳の品質管理に関する用語として、品質保証 (Quality Assurance) と品質評価 (Quality Assessment) がある。MQM Definition<sup>1</sup>によれば、品質保証は翻訳において誤りや欠陥を防ぐ方法を指しており、品質評価は翻訳が仕様に準拠している程度を指す。本稿では、これらの差異は取り扱わず、同一の観点から分析を行う<sup>2</sup>。

また、本稿では、代表的な QA スキームとして、具体的に以下に挙げた 3 つを分析対象とする。

- Translation services – Requirements for translation services (ISO 17100:2015) [3]
- Multidimensional Quality Metrics [4]
- JTF 翻訳品質評価ガイドライン [5]

以下、それぞれ ISO, MQM, JTF と言及する。これら 3 つを取り上げるのは、それぞれ、ISO は翻訳サービスを包括的に扱うものとしてメジャーなものであるため、MQM はより具体的で改造の高いものを提供しているため、JTF は日本語の翻訳を評価する際に用いることを想定しているためである。

## 3 評価視点について

翻訳品質の評価について、田中は、大半の企業が独自に作成した評価モデルによる評価項目を設けているこ

<sup>1</sup><http://www.qt21.eu/mqm-definition/definition-2015-06-16.html> を参照のこと。

<sup>2</sup>これは、品質評価と品質保証の差異が重視でないことを意味しない。翻訳プロセスをモデル化するにあたっては、プロダクトの評価である品質評価の記述をプロセスとしての品質保証の記述に書き換えることが必要であり、翻訳プロセスのモデルを構築するという点から、それらの差異に意識を払うことは意義がある。しかし、本稿は QA 記述を言語学的カテゴリーと対応付けることを目的としており、その場合にはむしろ品質評価と品質保証を同一の観点から分析する必要がある。なぜなら、言語学的カテゴリーを軸に品質評価と品質保証を同様に扱うことで、それらのより正確な同異、およびそれらの対応関係の把握につながるからである。それは、品質評価の記述を品質保証の記述に書き換えるために必要な過程である。

とからさまざまな評価項目が用いられていると述べている [6]。しかし、同時に、問題点の種類を表す用語として「正確さ (Accuracy)」、「流暢さ (Fluency)」、「用語 (Terminology)」、「スタイル (Style)」の 4 項目に関しては、ほぼすべての評価モデルで用いられていることを指摘している。本稿で分析対象とする ISO, MQM, JTF についても、すべてに共通する評価視点として上記の 4 項目が挙げられる。これらは、MQM Definition では、以下のように定義されている。

- Accuracy:  
The target text does not accurately reflect the source text, allowing for any differences authorized by specifications.
- Fluency:  
Issues related to the form or content of a text, irrespective as to whether it is a translation or not.
- Terminology:  
A term (domain-specific word) is translated with a term other than the one expected for the domain or otherwise specified.
- Style:  
The text has stylistic problems.

本研究でもこれら 4 項目について取り上げ、分析をする。本稿では、特に様々な言語学的カテゴリーとの対応が見られる「流暢さ」についての詳細を記す。

## 4 言語学的カテゴリーとの対応

### 4.1 「流暢さ」

以下、「流暢さ」について各 QA スキームの記述、および言語学的カテゴリーとの対応をまとめ、それぞれに説明を加える。

MQM において、「流暢さ」に関してはいくつもの下位分類がある<sup>3</sup>が、その下位分類の記述の言語学的カテゴリーとの対応は (1) の通りである。なお、ここでは、MQM Definition において、Fluency の子項目となっているものを挙げた<sup>4</sup>。

<sup>3</sup>各下位項目の定義は MQM Definition を参照のこと。

<sup>4</sup>MQM では、項目間の親子関係に不整合が見られる部分がある。MQM Definition において、例えば fluency の部分ではその子項目として duplication は記されていないが、duplication の部分ではその親項目として fluency が挙げられている。本稿では、fluency の部分で子項目として記述されている項目について記す。ただし、fluency の部分で子項目として style も挙げられているが、style は

### (1) MQM

- a. ambiguity: 語, 句, 文, 文章
- b. character-encoding: 文字
- c. coherence: 文章
- d. cohesion: 文章
- e. corpus-conformance: 語, 句, 文, 文章
- f. duplication: 記号, 語, 句, 文, 文章
- g. grammar: 語, 句, 文
- h. grammatical-register: 語 (, 句, 文)
- i. inconsistency: 語, 句, 文, 文章
- j. index-toc: 文書構成要素
- k. link-crossreference: 文書構成要素
- l. non-allowed-characters: 文字
- m. pattern-problem: 句読点
- n. sorting: 文書構成要素
- o. spelling: 語
- p. typography: 記号, 文字 (, 語, 句)
- q. unintelligible: /

(1a) は意味にかかわるものである。意味を持つ最小単位は形態素であるため、形態素以上のすべての言語単位が対象となっている<sup>5</sup>。(1b) は文字コードの問題であり、対象は文字のレベルである。一貫性、結束性は文章のレベルで決定されるものであるため、(1c) も (1d) も文章を対象とした記述である。ただし、(1c) に関して、その決定因は何らかの文や句、語といった単位となり得る。(1d) については、例えば、論理関係を示すための特定の語の不足などによって引き起こされることがあるため、その決定因となるものとしては、語や句なども挙げられる。(1e) はあらゆる言語単位を対象とした記述であると考えられる。(1g) は文法、および統語法に関するものであり、これは語、句、文を対象とした記述である。(1h) については、MQM は具体例として語に関するものを提示しているが、文法上のレジスターにかかわる問題は句や文の単位でも起こり得る。(1i) は理論的にはあらゆる単位で矛盾をきたすことがあり得ることから、語、句、文、文章が記述の対象となっている。(1j, 1k, 1n) は目次、リンクお

fluency と同じレベルの項目として扱うのが適切であると考えられるため、style は fluency の子項目としては扱わない。

<sup>5</sup>そのような意味では、(1a) には「形態素」も追加しておく必要があるが、実際に翻訳という行為について、その操作の対象となるのは一般的に所与のものとして考えられる「語」以上のカテゴリーであると考えられる。そのため、これ以降「語」以下のカテゴリーについては QA 記述の対象として扱わないこととする。

よびリストに関わるものであり、言語学的カテゴリーではない文書構成要素を対象とした記述である。(1l)は文字、記号を対象とした記述であるが、実際にこれを判断するためには、仕様書などの言語外要素を参照する必要がある。(1m)は句読点などの記号を対象とした記述である<sup>6</sup>。(1o)は、語の表記上の問題を扱っており、記述の対象は語である。(1p)は(1o)に含まれない表記上の問題を扱っていると言え、記述の対象は記号や文字であるが、その判定には語や句を参照しなければならない。(8f)は判読不能部分の特定ができない限り、対象を特定することはできない。

ISOから「流暢さ」に関する記述を抽出すると、具体的には syntax, spelling, lexical cohesion の3項目についての評価となっている。これらの記述の言語学的カテゴリーとの対応は(2)の通りとなる。

## (2) ISO

- a. syntax: 語, 句, 文
- b. spelling: 語
- c. lexical cohesion: 語(, 句, 文, 文章)

(2a)は統語法について記述したものであることから、記述の対象は語、句、文である。(2b)は語の表記の問題であるため、対象は語である。(2c)については、基本的に語を対象としているが、語の結束性は語単体では定義され得ず、必ず句や文、文章を参照する必要がある。

JTFのエラーカテゴリーには、概要レベルと詳細レベルの2レベルがあり、「流暢さ」に関する詳細レベルの記述と言語学的カテゴリーとの対応は(3)の通りとなる。

## (3) JTF

- a. 誤入力: 語, 記号
- b. 誤字: 語(, 記号)
- c. 同音異義語誤り: 語(, 文字)
- d. 文法誤り: 語, 句, 文
- e. 誤用: 語, 句, 文
- f. コロケーション誤り: 語(, 句, 文)
- g. 待遇表現誤り: 語, 句(, 文, 文章)
- h. 不整合: 語, 句, 文, 文章
- i. あいまい: 句, 文

<sup>6</sup>ただ、定義から考えれば、音素配列論的なパターンまで含まれることになり、spellingと本質的に区別できなくなる。実際には、あくまで記号(等)についての記述であると考えられる。

- j. 読解不能: /
- k. その他: /

(3a)は入力時の誤りについての記述であり、その対象となるのは語や記号であると判断できる。(3b)はスペルミスを含む誤字について記述しており、対象となる単位は基本的に語である。ただし、ここには句読点などの記号の誤りも含まれる。(3c)については、基本的に語のレベルでの問題であると考えられるが、単なる文字表記上の問題と本質的には区別できない。(3d)は文法上の誤りを扱うものであるため、記述の対象は語、句、文である。(3e)は意味に関するあらゆる誤用を含むものであり、意味を持つ最小単位である形態素以上の語、句、文を対象としている。(3f)はコロケーションについての記述であるため、その対象は語であるが、本質的にコロケーションに関して判定をする際は句や文を参照しなければならない。(3g)は待遇表現についての記述であることから、語、および句が記述の対象である。ただし、適切な待遇表現が用いられているか否かの判定には、文や文章を参照する必要がある。(3h)については、語、句、文、文章といったあらゆる単位で不整合性が確認され得る。(3i)は曖昧性についての記述である。曖昧性には語彙的なものと統語的なものがあるが、ここでは「係り受けなどが原因」となるような統語的曖昧性について記述されているため、その対象となるカテゴリーは句、文となる。(3j)については判読不能である部分の対象の特定はできない。また、(3k)についてもその他のあらゆる問題が含まれるため、対象の特定はできない。

## 4.2 「正確さ」、「用語」、「スタイル」

以下、「正確さ」、「用語」、「スタイル」について比較的具体的に記述されているMQMとJTFの記述をまとめ、特に言及が必要なものについて若干の説明を加える。

「正確さ」については、基本的に文章を対象とした記述であり、対応する言語学的カテゴリーは基本的には文章となると判断できる。例えば、MQM、JTFの「正確さ」についての、下位項目は(4, 5)の通りである。

### (4) MQM

- a. addition
- b. improper-exact-tm-match
- c. mistranslation

- d. omission
- e. over-translation
- f. under-translation
- g. untranslated

(5) JTF

- a. 誤訳
- b. 抜けと余分
- c. 未翻訳
- d. その他

「正確さ」に関する記述は、これらは基本的に文章を対象とした記述であるが、実際には (4a, 4d, 4e, 4f, 5b, 5c) はあらゆる言語単位に関して起こり得る。

「用語」については、すべての記述が専門用語、固有名詞に関するものであるため、言語学的カテゴリーとしては、語、ないしは句に対応する。MQM, JTF の「用語」についての下位項目は (6, 7) の通りである。

(6) MQM

- a. term-inconsistency
- b. termbase
- c. domain-terminology

(7) JTF

- a. 指定用語違反
- b. 特定分野用語違反
- c. 用語不統一
- d. その他

「スタイル」は基本的に文章全体から決まるものであるため、その記述は文章を対象としている。MQM, JTF の「スタイル」についての、下位項目は (8, 9) の通りである。

(8) MQM

- a. awkward
- b. company-style
- c. inconsistent-style
- d. register
- e. third-party-style
- f. unidiomatic

(9) JTF

- a. 指定スタイル違反

- b. 特定分野スタイル違反
- c. スタイル不統一
- d. その他

(8f) については、文章全体、および言語外の慣例から判断されるものの、直接的な対象となっているのは語句である。

## 5 おわりに

以上、本稿では、翻訳プロセスを高解像度で語るための第一歩として、ISO, MQM, JTF の 3 つの QA スキームを対象に、「正確さ」、「流暢さ」、「用語」、「スタイル」という 4 つの項目に関して各記述と言語学的カテゴリーとの対応についてを記述することで、QA の記述解像度を分析した。

翻訳プロセスについて解像度が高いモデルを構築していくために、今後は扱う QA スキーム、および評価視点を追加し、非言語学的カテゴリーも含め、詳細に分析していく必要がある。また、本稿をある種の叩き台として、実務翻訳の QA 記述について議論が惹起され、今後より活発化されることが期待される。

## 謝辞

本研究は JSPS 科研費 (課題番号: 19H05660, 19K23073) の助成を受けている。

## 引用文献

- [1] ミカエル・ウスティノフ. 翻訳: その歴史・理論・展望. 白水社, 2008. 服部雄一郎 (訳).
- [2] Kyo Kageura. The status of documents and related concepts in translation and in library science. In *Proceedings of the 9th Asia-Pacific Conference on Library and Information Education and Practice*, 2019.
- [3] International Organization for Standardization. Translation services – requirements for translation services (ISO 17100:2015), 2015.
- [4] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH and QTLaunchPadn. Multidimensional quality metrics, 2015.
- [5] 日本翻訳連盟. JTF 翻訳品質評価ガイドライン, 2018.
- [6] 田中千鶴香. 実務翻訳における翻訳品質評価プロセス. 言語処理学会第 23 回年次大会発表論文集, pp. 489–492, 2017.