

改めて、翻訳とは何か：Google NMTが使える時代に

影浦峽 (東京大学)

1 はじめに

Google 英日翻訳が NMT を採用して、目標言語の流暢さは格段に向上した。2016年11月以来ネットを見ると、Google 翻訳の精度向上で翻訳業者の未来は暗いのではとか¹、論文執筆や抄録会に使おうとか²、英語を勉強する必要がなくなるのではとか³、様々な見解が見られる。多くは印象論的なものであるが、それらは一般の人々の、あるいは言語学者や機械翻訳研究者の、さらには一部の「翻訳者」の、「翻訳」理解を反映してもいそうである。一方、どんなに Google 翻訳の精度が向上しても、ハイエンドの翻訳は決して人間のようにできるようにはならないという主張もある。その理由付けにしばしば持ち出されるのが、意味や文化である。本稿の目的は、強引に単純化したものの必ずしも外的外れではないこの議論の配置の中で取り残されている領域が存在すること、そのうち一定程度は、計算機処理で扱えるかたちで問題を定式化することができる可能性があることを指摘することにある。本稿では問題提起的を行うにとどまり、体系的に定式化の方向性を提示するわけではない。なお、本稿では、言語の創造に関わる翻訳は考えない⁴。また、翻訳に対する説明をめぐる問題も⁵、社会的責任をめぐる問題も、そのものとしては扱わない。

2 Google 翻訳ができないこと

今のところ、Google 翻訳を始めとする MT には例えば以下の文章の翻訳は「原理的に」できない。

¹Speedque01 「Google 翻訳の精度が半端ない！翻訳業者の未来はどうなる」(ママ) <http://www.outward-matrix.com/entry/2016/11/17/070000>

²cosine 「Google 翻訳の精度が飛躍的に向上！～その活用法を考える～」(ママ) <http://www.chemstation.com/blog/2016/11/translation.html>

³Yuki Sako 「Google 翻訳、ニューラルネットワークが導入されて精度が大幅上昇」<http://www.yukisako.xyz/entry/google-translate>

⁴これについては、影浦 (2015) を参照。また、森安 (2002) のルターとドイツ語への言及、Benjamin (1923)、魯迅や二葉亭四迷の翻訳論は参考になる。

⁵ML との関係では Lei et al. (2016) が関連する問題を扱っている。

1. “Human” consists of five letters.
⇒ 「人間」は5文字で構成されています。

2. States Parties shall assure to everyone within their jurisdiction effective protection and remedies, through the competent national tribunals and other State institutions, against any acts of racial discrimination which violate his human rights and fundamental freedoms contrary to this Convention, as well as the right to seek from such tribunals just and adequate reparation or satisfaction for any damage suffered as a result of such discrimination.

⇒ 締約国は、管轄内のすべての人に、有能な国内裁判所及びその他の国家機関を通じて、人権及び本条約に反する基本的自由を侵害するあらゆる人種差別行為並びに捜索する権利そのような差別の結果として被った損害に対する公正かつ適切な賠償または満足をそのような法廷から得ることができます。

1 は中世哲学における「質量代表」で、現代言語学も計算言語学も基本的にこうしたことを捨象しているので、できないのは当然である。一方、人間の翻訳者は基本的に困難なく扱える⁶。

2 は人間の「翻訳者」でもワークショップ等で問うとできない人が少なくない。できないのは文の構造が複雑だからではない(その意味では人間も Google 翻訳もある程度はできる)。多くの言葉自慢の自称「翻訳者」も Google 翻訳も、この文を「英文和訳」しようとする。けれども、プロの翻訳者ならば、この文は人種差別撤廃条約という歴史社会的に固有の役割を担う文書の一部で、日本政府はヘイトスピーチの部分に保留をつけているものの同条約を批准しており従って日本語版正訳あるいは仮訳が存在するはずだから他に(例えば子ども向けの翻訳であるといった)指定がなければ、公式の訳を用いるという判断をしなくてはならない。この文に対して、「流暢でわかりやすい訳文」を作ることができても翻訳にはならな

⁶この問題が瑣末なものに見るとするとそれは「言語学的」的に言語を見るからであり、言語の実態的作用の観点からはこれは本質に関わっている。

い。この問題を扱うメカニズムもまた、Google 翻訳にはそもそも組み込まれていない（この例の場合は言葉自慢の X 文 Y 訳家にも組み込まれていない）。ちなみに、単位は異なるが、この問題は固有名の扱いの問題に近い。実際、Google 翻訳では固有名も原理的に扱えていない⁷。

3 X 文 Y 訳と翻訳

これらの例が理論的に指し示す「翻訳」という行為を、言語表現という存在の性格付けの観点から整理しておこう。

単純化して言うと、言語学も機械翻訳も、有限の規則と語彙に基づき、潜在的に無限の適格文を生成したり処理したり変換することを言語及び言語の操作と見なす。「X 文 Y 訳」はその枠組みでの言語間変換操作である。この枠組みでは、与えられた起点言語の文書は、たまたま与えられた、他の何であつてもよかつたものとみなされる。これに対して、翻訳は、与えられた起点言語の文書を、現実に社会の中で他でもなくそれであるものと見なした上で、それに対して加えられる行為である。必然ではないが、言語表現の単位としては前者が文を、後者が文章を基本としがちなのは、そのためである。

次のような Foucault (1968) の言葉は、言語学／機械翻訳と翻訳の相違を明確に示すとともに、それがより広い範囲で観察されるものであることを示している。

言語 [ラング] とは、無限数個の言語運用 [パフォーマンス] をゆるす有限な規則の集合であるのだ。これに対して、言説とは、じっさいに述べられた言語的まとまり [シークエンス] のみからなる、つねに有限で現働的 [アクチュアル] に限定された集合である。[...] 言語の分析が、ある言説の事実に関して問う問いとはつねに、「どのような規則にもとづいて、このような言表がつくられたのか？ またしたがって、どのような規則によって他の同じような言表をつくることのできるのか？」であるおんい対して、言説の記述が立てるのはまったく別の問いであつて、「このような言表が出現した、しかも、他のいかなる言表もその代

わりには出現しなかつたのは、どのようなわけなのか？」という問いなのだ。

言語処理も言語学も X 文 Y 訳も（言語学と X 文 Y 訳はほぼ定義上、言語処理は事実上）「言語の分析」（あるいは研究）の側にあるのに対し、翻訳はむしろ「言説の記述」（翻訳の場合は言説への介入と操作）の側にある。実は我々の言語表現行動も、論文執筆活動を考えればわかるように、基本的に後者に属する。X 文 Y 訳が文単位で言語学的な範囲に矮小化した意味の等価性を基準とするのに対し、翻訳に求められるのは、起点言語の文書に関与する言語表現域に対応する目標言語の言語表現域において、起点言語の文書の位置づけに対応する目標言語の位置を与えることである⁸。

そう考えるならば、データに基づくアプローチは、現実に存在する、それ以外ではなかつた言語表現を、他でもありえたまたまのデータとみなし、他でもありうるものに適用させる規則の抽出のみに使つて、そのものであるという存在を落とすという、いわばもつたいないことをやっていることになる。

時間の問題を別にすると、せいぜい棚延長 50 メートル程度しか読んでいない人間に Google 翻訳よりも優れた翻訳ができるのは、そもそも「翻訳」という名前で行っている行為が両者で異なるからである。X 文 Y 訳や Google 翻訳に人間のようなハイレベルな翻訳ができないのは、そのような意味での翻訳をやっていないのだから当然である。では、人間が翻訳としてやっている行為は、機械的な処理に落とせないのだろうか。次に、消極的および積極的な側面から可能性と方向性を考えることにしよう。

4 意味と文化

コンピュータは意味がわからないから（筆者もそう思つてはいる）、あるいは文化の機微がわからないから、上質な翻訳はできないといった主張を検討しておこう。基本的にこうした主張は、第一に「言語」の操作域を言語学や X 文 Y 訳が扱う範囲にまず限定し（その部分は説明的言語化や形式化が可能だとし）た上で、それ以外の部分については操作可能なレベルまで説明的言語化も形式化もしようとせず人間にしかできないというかたちでなされることが多い。

意味について、意味論は「意味というものが存在し我々はそれについては（意味がわかるという

⁷多くの固有名は実際にはとても正確に翻訳されるが、“Tokyo University of Technology has many students.”と“Tokyo Institute of Technology is not Tokyo University of Technology.”を日本語に訳してみると「原理的に」扱えていないことがわかる

⁸注や解説は目標言語の言語表現域が対応付けに不十分である場合にそれを補完するメカニズムと見なすことができる。

意味で) わかっている」という前提でなされるし、意味理解・認知の問題も意味も意味を理解することも我々はわかっているという前提でなされる⁹。ところが「それはいかなることか」はわかっていない。視点を変え、意味がわからないとき人はどうするかを見ると、典型的には辞書を引くが、辞書には意味ではなく言葉が書いてあるだけである。

文化の問題は機械翻訳の問題だけでなく、人間の翻訳が不可能であるとの主張にも持ち出されることがある。翻訳の不可能性は神学的主張でないならば論理的に翻訳の理念的可能性を要請せざるを得ないし、文化の差異を実体的に存在するとするならば、一方では「同じ言語」性も解体するところまで至る歯止めは存在しなくなるし(誰もが違う文化を持っている)、他方では言語のレベルで捉えられないものならば言語の操作において考慮する必要はないものとなる。

ここから、「意味」や「文化」と言われるものを、表出され記録された言語表現において捉えることはできないのかという問いが出る。

5 関与言語操作域・表現域

実際、言語処理における意味の分散表現は、言葉の意味をそれなりに捉えていると言ってよい。翻訳との関係で言うならば、問題はむしろ、そこで捉えられているのが、言語学的なかたちに矮小化された「意味」に過ぎないということであろう。これに対して、意味が言語表現の形式と言語表現の社会的・歴史的位置づけの問題に還元される事例と、対象文書に関与する文書集合の選択の問題として現れる事例をここでは考える。

前者の例として、2節の事例2そして固有名がある。これらの言語表現の操作は、その表現の社会的なその位置づけを反映した言語表現集合上の位置づけに対応しており、言語学的な意味での文脈や意味には依存しないので、原則としてすべての文脈で同じ言語表現を参照することが求められる。別の言い方をすると、これらの表現には言語学が言うところの意味ではなく、法的な意味の透過性を維持することが要求され、それは固定された表現の利用によって実現される。逆説的に(とはいえ表現形式がなければ言語における意味はないのだから当然で)、意味の詳細な扱いは表現形式の問題に還元される。

もう一つ、いわゆる「文化」に関わる翻訳の例を考えてみよう。多和田葉子がドイツで朗読会を

したとき、多和田の詩の「畳」がドイツ語で翻訳者により「絨毯」と訳されていたのに憤慨したある日本人がに対し、多和田は、ドイツ語で「Tatami」とすると若いおしゃれなカップルの生活をイメージしてしまい、日本の「畳」に対応しない、むしろドイツで対応するのは「絨毯」だと、「絨毯」訳を擁護している(多和田, 2006)。この事例は、「文化」という言葉で示されるような言語表現の属性に従って、起点言語文書に関与する言語表現域を確定し、それに対応する目標言語での関与言語表現域を確定し、その中で対応を求めれば表現のレベルで処理できそうである。

前者の問題は、文学の領域では引用としてよく知られており(Compagnon, 2010)、言語表現を以下のように扱うことがポイントである。

- 言語表現を歴史的に固有のものに見なす。
- 言語表現の形式を参照する。

「畳」の例も、関与する言語表現域の確定は当該の詩と潜在的引用関係を結ぶ領域の確定と見なすことができるし、質量代表の問題は文脈に依存しない引用の課題と見なすことができる。実際のところ、論文を書くことを考えてみればわかるように、我々の言語表現行為において、これらは当たり前前に実践されていることであり¹⁰、科学の信頼性も「合理的に疑問の余地なき」という近代社会の原理的規範も、これを基盤としている。

6 言語表現の階層

前節末尾の「引用」について、人種差別撤廃条約であれば(他の指定がない限り)正訳を参照しなくてはならないが、より一般的な文書であれば、既訳があったとしても、それを必ず参照しなくてはならないわけではない。このことは、翻訳においては言語表現を歴史的に固有のものに見なす必要があるとともに、その上で、固有の言語表現に、「階層」を認定する作業が伴うということでもある。

固有の文書に対する「階層」は、論理的、歴史のおよび/あるいは社会的に、また、大域的あるいは局所的に認定される。例えば、目標言語側の憲法が起点言語の文書で起点言語で引用されているとき、立憲主義の社会においては、目標言語側の憲法の文言そのものが復元される必要がある。一般に、法の支配を原則とする社会では、法は、社会的に言語表現レベルでの固定性が高い、上位の階層に属する。

¹⁰ 言語処理では剽窃検出がこれを扱う。

⁹ 私たちが認知的行為として行っていることに気づいてさえいない行為の認知プロセスを対象化して問うことは論理的に不可能なので、認知の研究は必然的にこのようになる。

局所的には、例えばマニュアルの翻訳において、過去のマニュアルとの一貫性が求められる場合は多い。実際のところ、翻訳メモリはそれに対応しているが、一般に、クライアント側に言語表現の固有性に対する認識がないため、翻訳メモリの来歴を含めたデータ管理は、翻訳者の専門的な技能として評価されていない。

ちなみに、言語表現の固有性における階層の問題は、研究活動において当たり前実践されている。誰の何を引用するか、ある技術については、原則として元々の提案者の文献を引用するといったことである。前節で述べたように、科学の暫定的妥当性も社会における「合理的に疑問の余地なき」という概念も、これに依存する。剽窃検知はこのような文書の階層性が前提とされているからこそタスクとして定義されるが、機械翻訳を含む言語処理一般では言語表現のこの側面は、上で示唆したように技術的に操作可能な問いとして定式化できる範囲はかなりありそうであるにもかかわらず、あまり考慮されていない。

7 おわりに：等価性の義務レベル

本稿で素描してきたことは、翻訳における等価性の観点から見ることもできる。いわゆる「X文Y訳」や機械翻訳では、主に言語学が捉える範囲での言語単位である文を対象に、言語学が捉える範囲での「意味」の等価性を保つことが目的となっている。

これに対して、別レベルの等価性があり、それらは、言語表現の固有性とその階層の操作のレベルに関わる。例えば「文化」は、言語表現域をそれに応じて確定した中で認識される等価性と見なすことができ、固有な表現域に依存しない形式の等価性と見ることができる。このように考えるならば、翻訳者が文化や意味を考慮すると称して行ってきたことは、ある言語表現を人間が生成するときには常に関与する言語表現域が想定されているという、当たり前の性格を復元することに他ならない。従って、「ハイエンドな」翻訳の課題は、少なくともこれまでよりは広い範囲で言語表現の操作に還元できるはずである。

与えられた文書に関与する言語表現域を固有のものとして確定し、その中で等価性を維持して表現を確定しなくてはならないという翻訳の要請、より一般に、妥当な言語表現のためにはそれを支える歴史的・社会的・論理的に固有の言語表現域を参照しなくてはならないという言語表現一般の要請は、議員が脱法を促す発言をするような社会

では¹¹、あまり流行らないかもしれないが、言語表現の変換において言語表現の性質に応じて維持されるべき等価性の概念が不要であるならば、狭義のX文Y訳でふつうの意味の等価性を保つことさえ不要になるし（どんな言語表現であれ好きなように解釈すればよい）、従って結局のところ翻訳はいらなくなる。

「ポスト真実」と称される時代に、人間の側が、翻訳という行為にも反映されている言語表現活動の基本的要請をますますないがしろにしつつあるときに、機械翻訳の側で、翻訳に求められる言語表現の等価性を、より公共的な言語表現という近代の理念に則したかたちで扱う準備ができたというのは、皮肉なことではある。

謝辞

本研究は、日本学術振興会科学研究費補助金基盤(A)「翻訳知のアーカイブ化を利用した協調・学習促進型翻訳支援プラットフォームの構築」(研究課題番号:25240051)の支援を受けて行われた。本稿の一部は『自然言語処理』第19巻2号に掲載された筆者による巻頭言「翻訳を通して『言葉』を見ると?」を敷衍したものとなっている。

参考文献

- Benjamin, Walter. (1923/1968) "The task of the translator," (Zohn, H. trans.)
- Compagnon, Antoine. (1979) *La seconde main ou, le travail de la citation*. Paris: Seuil, 1979 (今井勉訳『第二の手、または引用の作業』水声社)
- Foucault, Michel. (1968a). Sur l'Archéologie des Sciences: Réponse au Cercle d'Épistémologie. *Cahiers pour l'Analyse*, 9 (Généalogie des Sciences), p. 9–40. (石田英敬訳「科学の考古学について—認識論サークルに答える」)
- 影浦峽 (2015) 「翻訳の社会的意味」『ひつじ意味論講座7 意味の社会性』ひつじ書房.
- Lei, Tao et al. (2016) "Rationalizing neural predictions," EMNLP2016.
- 森安達也 (2002) 『近代国家とキリスト教』平凡社ライブラリー.
- 多和田葉子 (2006) 「ある翻訳家への手紙」『翻訳家の仕事』岩波新書. p. 169–174.

¹¹例えば日本維新の会のあだち康史議員は2017年1月16日、twitter上で「環境基準を厳格に守ること自体がナンセンス、有害でさえある」と、法令無視を推奨する発言をしている。