

## 機械翻訳は言語帝国主義を終わらせるのか？： そのしくみから考えてみる

西島 佑

上智大学大学院グローバル・スタディーズ研究科特別研究員

本稿は、未来において高度な機械翻訳が実現すると仮定した上で、それが英語の一極状況を終わらせるのかどうか、そのしくみから考察している。機械翻訳には、英語の一極状況を終わらせるのではないかという期待がよせられているが、本稿でそうしたことは起こりえないと結論している。機械翻訳の現代のパラダイムは、「ニューラル機械翻訳」とよばれており、人間の言語情報を学習していくことでつくられている。これは言語のデータが大量にある大言語間であれば、英語の影響力を低下させるだろうが、言語のデータが乏しい少数言語の場合、英語の影響は依然として残ると予想される。

キーワード：機械翻訳 言語帝国主義 国際共通語 英語 社会と技術

### 1. はじめに

国際共通語というと、おそらく多くの人々が思いうかべるのは英語だろう。英語は、グローバル化の言語的側面であり、現在では国際共通語とよべる機能をもっている。しかしながら、この国際共通語としての英語の地位には批判もある。

わかりやすい批判の一例としてあげられるのは、英語ネイティブと非ネイティブの非対称性だろう。たとえば英語非ネイティブの研究者が英語で論文を執筆する際、通常ネイティブチェックが要求される。だが英語ネイティブにその必要はない。くわえるとネイティブチェックには、金銭的やりとりが一般に生じるが、ここでも英語ネイティブ使用者は金銭的受益者となる一方で、非ネイティブ使用者は支払い者となる。

また多くの人々が国際共通語である英語を通じて情報を受容することは、英語圏で議論される情報の流通性だけが高まるという指摘もある。つまり英語圏であるアメリカやイギリスなどの事情は、他言語にも通じやすいが、英語以外の言語で論じられている情報は相対的に伝わりにくくなっている。

このほかにも国際共通語としての英語については、すべての人々がみにつけることの非

現実性、学習コストの問題、ポストコロニアルな関係、英語だけが価値ある言語とみなされるなどといった批判がある（詳細については、たとえば木村（2016））。もちろん英語という言葉そのものに問題があるわけではない。人々が英語という言葉そのものに愛着をもつのは自由である。ただし国際共通語としての地位にある英語が、ほかの言語にはない特権性をもってしまふことは否定しがたい。

こうした国際共通語としての英語の覇権的な地位については、すでに研究者らも注目しており、「言語帝国主義」（たとえば Phillipson (1992); 三浦信孝・糟谷（編）(2000); Crystal (1997=2003); サンダルソラ(2010))という概念で論じられている。そのうえで言語帝国主義とどうむきあうのかについてはさまざまに議論されてきた。本稿でも言語帝国主義とどうむきあうのかというテーマを論じたい。ただし、これまでの議論とは、とりあげる観点が異なっている。本稿でとりあげるのは、「機械翻訳」というテーマである。

## 1.1 機械翻訳

機械翻訳とは、たとえば『Google 翻訳』のようなものをいう。定義を述べると、「機械翻訳とは、コンピューターのような知的機構を利用した自動翻訳」となる<sup>1</sup>。

機械翻訳というと、誤訳ばかりで使いものにならないといった印象をもつ方もいるかもしれない。たしかに以前は、機械翻訳の研究者でさえ、その精度がなかなかあがらないことを認めていた（長尾 1986, 第 8 章）。ところが近年この機械翻訳は、急速に技術発展し、社会のさまざまな側面で活用されるようになってきている。

たとえば公文書の多言語翻訳を政策として実施している EU では、その作業に機械翻訳が不可欠となっている（長尾 2016, 1）。日本でも 2014 年に総務省が「グローバルコミュニケーション計画」を発表し、2020 年東京オリンピックをめどに、「言葉の壁」をなくすことがかけられ<sup>2</sup>、そのために機械翻訳のアプリケーション『Voice Tra』が国立行政法人の情報通信研究機構によって開発されている。科学技術文献のデータベース JDream III には、JSTchina という中国語文献の抄録を日本語で読めるサービスがあるが、近年では文献数の増加にともない、一部機械翻訳のみによって翻訳が行われているようである（中澤 2017, 305）。

このように機械翻訳は、少し前に考えられていたようにまったく役にたたないものとはいいがたくなりつつある。今後、機械翻訳が技術的により発展し、観光地や国際会議等で使

---

<sup>1</sup> ただし機械翻訳の定義は、機械翻訳がどの程度実用化されているのかによって変化していく。たとえば新田は、現状では機械翻訳は人間の翻訳のかわりになっていないことから、機械翻訳をコンピューターのような知的機構を利用して、人間の翻訳を援助する方法・理論・手段と定義している（新田 2012, 24-5）。現段階では、新田の定義のほうが正確だろう。

<sup>2</sup> 以下の総務省のホームページから確認できる。

[http://www.soumu.go.jp/main\\_content/000285578.pdf](http://www.soumu.go.jp/main_content/000285578.pdf) (最終アクセス 2018 年 5 月 10 日)

用できるようになれば異言語間での対話の心理的なハードルがさがるとも考えられる。

国際共通語、あるいはリンガフランカの歴史を研究した Ostler は、著書 *The Last Lingua Franca: English Until the Return of Babel* のなかで、機械翻訳が高度に技術発展することで、今後人類が国際共通語を必要としなくなり、英語が最後のリンガフランカになるだろうと述べている (Ostler 2012, 281-3)。もしそのような事態となれば、機械翻訳は、言語帝国主義とむきあう議論においても大きなインパクトとなるかもしれない。

Ostler の予想はあたるのだろうか。筆者自身は、現代の機械翻訳のパラダイムのもとでは不可能であり、むしろ言語帝国主義とされる事態がより複雑に変化していくと考えている。とはいえ、機械翻訳という技術によって今後どのように人間社会のあり方が変わりうるのか、本稿の関心からいって言語帝国主義とされる状況がどう変化するのかを考察してもよい段階になってきているのは否めない。

そこで未来において高度な機械翻訳が登場すると仮定したうえで、本当に機械翻訳という技術が、言語帝国主義として語られる人間社会の現状を変えうるのか考察してみたい。

## 1.2 構成

構成としては、まず次の第 2 節で従来の言語帝国主義とむきあってきた議論に対して、機械翻訳がどういう点で独自性があるといえるのかを考えてみたい。第 3 節では、機械翻訳が言語帝国主義とされる事態を変えることができるのかを考察するために、そのしくみについて簡単に述べる。しくみを理解することで機械翻訳には、なにが可能で、なにが不可能なのかがわかり、上記の問いにもこたえることができるからである。そして現代の機械翻訳のしくみでは、人間の言語状況の影響をうけることから、依然として諸言語を媒介する国際共通語のような機能が必要とされ続けることを指摘する。つまり言語帝国主義は、機械翻訳によっては終わらないことを述べる。そして第 4 節では、機械翻訳は、言語帝国主義を終わらせるわけではなく、むしろそのあり方を潜在化させていく旨を論じてみたい。

## 2. 先行研究のなかで機械翻訳はどう位置づけられるのか？

本節では、従来の言語帝国主義とむきあう議論をふりかえりながら、機械翻訳の独自性というものがなにかを考えてみたい。最初に研究者のなかでとくに注目をあつめている議論を記述し、次にそれらに対して機械翻訳がどういった点で独自といえるのかを述べる。

### 2.1 Van Parijs の提案

最初にとりあげるのは、ベルギーの哲学者 Van Parijs の提案である。Van Parijs は、ベーシックインカムを発案者としても知られている (以下、Van Parijs (2012)による; なお Van Parijs の議論については、木村 (2013)の要約が簡明である)。

Van Parijs は、まずグローバルな諸問題を議論するためには国際共通語が不可欠と訴える。たとえば環境問題のように、現代では1つの国家や地域のなかで解決できないものがある。そしてこの問題に対処するためには、さまざまな人々は国境をこえた議論の場に参加することがのぞましい。だが世界には、多くの言語があり、対話は容易ではない。ゆえに国際共通語としての英語が不可欠というわけだ。

それでは Van Parijs は、なにを問題としたのだろうか。彼は、効用理論を用いながら、英語圏の国々とそれ以外の国々とは、英語を学習するコストにおいて、大きな非対称性が厳然としてあると論じる。英語圏の人々は、第一言語として英語を学習できる環境にあり、異言語間でも英語を使用することができる。だが非英語圏の人々は、多くの場合英語を第二言語として学ばなければならない、そのためのコストも自分たちで払わなければならない。Van Parijs が問題としたのは、こうした英語圏とそれ以外の諸国とのあいだの英語学習コストの非対称性である。

この問題を解決するために Van Parijs は、英語を今よりも世界的に徹底すべきという大胆な主張を行った。英語を世界的に徹底するとは、文字通り地球上の全人類が英語を第一言語としてみにつけるべきだという提案である。この議論によると、最終的に全人類が英語を第一言語として使用するようになることで、学習コストの非対称性の問題を解決できるとしている。

Van Parijs は、そのための方策として、超国家的な政策的処方箋が必要だとも述べている。具体的には、英語圏の諸国は、英語によってさまざまな便益をうけていることから、英語税なるグローバル・タックスをはらうべきであり、その税収は非英語圏の英語教育につかわれるのが望ましいと述べている。

こうした提案は、英語を功利主義的に普及させることでネイティブと非ネイティブの差を縮める議論としては、もっとも体系的とすることができる。だが Van Parijs の議論は、グローバル・タックス的な英語税という点は置いておくとしても、英語の世界的普及を徹底するため、英語以外の言語をなくすべきと論じていることには批判が強い(たとえば、Schutter, Robichaud (eds.) (2016))。こうした議論が、英語以外の言語に慣れ親しんでいる人々にうけいられるとは考えにくい。

## 2.2 国際英語論

次に国際英語(English as a Lingua Franca)の議論について述べる。国際英語論とは、国際共通語としての英語を特定の社会集団(いわゆる「アメリカ人」や「イギリス人」)が所有する言語とは考えないようにすべきという提案である<sup>3</sup>。

---

<sup>3</sup> 国際英語論とされるこの議論は、歴史的に World Englishes(Kachru (1976))、English as an International Language(Smith (1983))、English as a Lingua Franca(Seidlhofer (2001); Jenkins (2007))などによばれてきており、それぞれ内容が異なっている(くわし

たとえばアジア圏で英語が使用される局面のことを考えてみよう。アジア圏で使用される英語とは、多くの場合、英語を第二言語として学んだ人々によるものである。こうした英語非ネイティブ間で使用されるとき英語は、ネイティブのようなイディオムや難解ないまわしを回避する傾向があり、平易な方がいい方が好まれ、「くずれた英語」であっても問題とはされないことが多々あり、彼らの第一言語の影響がみられる特徴がある。こうした英語は、もはやネイティブ英語とは別物といえる。そこで、ネイティブ英語とは別に、非ネイティブ間で使用されるような英語こそ、国際共通語として認知すべきとの提案がなされたのである。

この議論によると、国際英語は平易であり、ネイティブが使用する英語とくらべると、人々は相対的に習得・使用しやすい。また国際英語論によると、第二言語としてネイティブ英語を使用するとなると、人々は躊躇する傾向があるが、それが国際英語であればそうした心理的負担は軽減するという。その意味で国際英語は、非ネイティブにとって「自主的な英語」だとしている。アメリカやイギリスのネイティブ英語が国際共通語になると、特定の言語が覇権的な地位につく言語帝国主義となってしまうが、国際英語が国際共通語となれば、ネイティブと非ネイティブの非対称性が薄れるので、言語帝国主義も緩和できるのではないか、このように国際英語論では考えられている。

国際英語論の最大の強みは、すでにある程度現実となっていることだろう。非ネイティブ間で英語が使用されることは、現代では比較的良好であることといってもさしつかえはない。国際英語論は、提案としての実現性が高い。

この国際英語論につきまとう批判としては、本当にネイティブ英語の影響を無視することができるのかというものがある(Fiedler 2010, 9)。国際英語がつかわれる場が成立しても、その場にネイティブ英語の使用者があらわれると、やはり他の者もひきずられてしまうのではないか。そうした疑念がぬぐえないというのである。こうした批判は、とくに次で述べるエスペラント主義とでもよべる主張を行う人々から提起されている。

### 2.3 エスペラント主義

エスペラント語は、19世紀にポーランドの医師ルドヴィコ・ザメンホフによって構想されたことを起源とする。もともとエスペラント語は、思想的には諸民族の争いをとめるために考案された。しかし、20世紀をとおして次第にザメンホフの影響から離れ、エスペラント語の使用者たちによってこれを国際共通語にしようという動きとなり、広まっていった経緯がある。本稿では、こうしたエスペラント語を国際共通語にすべきという議論を「エスペラント主義」としたい。

---

くは(吉川 2016, 8-13)。本論で述べているのは、この内の English as a Lingua Franca の概念となっている。

現代のエスペラント主義者たちの主張を要約すれば次のようになる<sup>4</sup>。まずエスペラント語は、不規則変化といった文法的例外事項がなく、言語としては習得しやすい。またエスペラント語は、Van Parijs のように、諸言語をなくすことを目的とはしていない。この言語は、さまざまな土着語と共存することを目的としている。そしてエスペラント語は、第一言語使用者がほとんどいない言語であることをもって、国際共通語として望ましい。

仮に、エスペラント語が国際共通語となれば、大多数の人々は第二言語として学ぶことになるだろう。たしかにそうなれば、言語帝国主義として提起されているネイティブ・非ネイティブ問題や、英語のポストコロニアル問題等は発生しないか、ひかえめにいっても生じにくくなるだろう。

ただしエスペラント主義には、実現性が低いという批判は多い。特定の大国の言語でないことは、逆に世界的に普及させていく手段が乏しいともいえる（詳細は(カルヴェ 2010, 第18章)。エスペラント主義のこうした点をかんがみて、岡崎と鎌田は、エスペラント語と国際英語論の双方を評価しつつ、国際共通語としての英語を可能な限りエスペラント化(計画言語化)することを提案している(岡崎・鎌田(2016))。これは実現性に強みはあるがネイティブ英語の影響がぬぐえない国際英語論に、計画言語化という発想をつけくわえることで、両者の長所を生かし短所を補おうとする考え方といえる<sup>5</sup>。

## 2.4 機械翻訳はどう位置づけられるか？

それでは言語帝国主義とむきあう従来の議論のなかで、機械翻訳はどういった独自性があるのだろうか<sup>6</sup>。これまでの議論は、いずれもさまざまな言語の膨大なくみあわせを縮減する「国際共通語の機能」そのものを否定することはできなかった。これについて、少し説明が必要になるだろう。

さまざまな言語があるなかで、国際共通語の機能を否定できなかった理由は以下から説明することができる。仮に国際共通語の機能がないと考えてみよう。世界の言語の数が  $n$  個

---

<sup>4</sup> エスペラント主義の代表性をあらわすものとしては、1996年の第81回世界エスペラント大会で採択された「MANIFESTO DE PRAGO de la Movado por la Internacia Lingvo ESPERANTO」(国際語エスペラント運動に関するプラハ宣言)がある。

<sup>5</sup> このほかにも、たとえば木村は、個人が場面ごとに言語をつかいはけることで、英語への依存を減らす「節英」という概念を提唱している(木村(2016))。節英は、土着語の重視と複言語主義を多層的につかいはける方針である。岡崎や鎌田もそうだが、近年では、なにか1つの方策によって言語帝国主義とむきあうことよりも、複数の方策から対応策が考察されることが多い。

<sup>6</sup> 実現性についてもここで少しだけふれておきたい。先行する研究のなかでもっとも実現性が高いのは、すでにある程度現実となっている国際英語論の議論だろう。他方、高度な機械翻訳が実現するかどうかは、その基盤にある数学の手法(論理、確率、統計)が、言語を表現できるモデルをつくれるか否かによっている。ただし実現性は重要な論点だが、本稿ではひとまずカッコにいれておきたい。

(L1, L2, ..., Ln)だとすると、異言語間の対話(Li⇔Lj)には、Ln(n-1)の組みあわせがある。たとえば言語の数が 5000 だとすると、5000\*4999 となる。この組みあわせは膨大であり、事実上あらゆる言語を超えた対話が不可能であることを示している。だが、ここで国際共通語があるとすると、2n 通りの組みあわせに縮減される。この組みあわせを縮減する機能があることが、国際共通語を必要とさせているのだ。

先行する議論は、国際共通語の機能は不可欠としたうえで、その問題点を各々にとらえ、英語の世界的な普及 (Van Parijs)、非ネイティブ化 (国際英語論)、代替として計画言語の提案 (エスペラント語)、そして国際英語の計画言語化 (岡崎・鎌田) といった解決策を提唱してきた。

これに対して、機械翻訳の独自性とは、国際共通語の機能そのものをなくせると主張していることだろう。これは冒頭で述べた Ostler が、いづれなくなるという意味で英語を「最後のリングフランカ」としていることからもうかがえる。「機械翻訳が言語帝国主義を終わらせる」という主張が意味しているのは、言語の膨大な組みあわせを、国際共通語の機能なしに、技術的に解消できると信じていることにある。

そうであるならば、機械翻訳の手法が、そのようなことを可能とするしくみとなっているのかを見極める必要がある。そこで次節では、機械翻訳のしくみをみていくことで、そうした疑問に答えることをこころみたい。

### 3. 機械翻訳のしくみ

本節では、現代の機械翻訳のしくみでは、言語帝国主義は終わらない旨を論じたい。現代の機械翻訳は、人間の言語情報を統計・確率的に学習することでつくられているが、言語情報はすべての言語に均等にあるわけではない。そのため異言語間を媒介する機能を依然として必要とし続けると考えられる。本題にはいる前に、まずは機械翻訳のしくみを歴史と手法のパラダイムから簡単にみていこう。

#### 3.1 機械翻訳の略史と手法

機械翻訳は、人工知能の一種であり、実際に人工知能の歴史と共に発展してきたといえる (人工知能の歴史については、小林 (2015); 松尾 (2015))。機械翻訳というアイデアが生まれたのは、Warren Weaver というアメリカの数学者が、私的な手紙のなかで後の機械翻訳に通じる発想を述べたことから始まるとされる (中澤 2017, 299)。研究としての機械翻訳は、1954 年にジョージタウン大学の研究グループと IBM 社が共同で発表したことから始まっている。しかしながら、こうしてはじまった機械翻訳研究は、当時は衝撃をもって受け入れられたものの、その後遅々として進まない状況が続くこととなった (渡辺他 2014, 2)。

#### ・ルールベースの登場

1970-80年代の機械翻訳は、ルールベース(rule base)とよばれる手法が主流であった。ルールベースとは、人間が文法をてがかりに言語を処理するルールを設計し、コンピューターのような知的機構によって自動的に翻訳がおこなわれる手法である。人間が文法書や単語帳から言語を学習するように、文法と辞書的なデータを用いて機械翻訳をつくればうまくいくに違いないというわけだ。

ところが、この手法には問題が生じた。問題とは、文法には多くの例外事項があり、そのような例外事項を人間が逐一設計していくのは現実的ではないことであった。人工智能がさまざまな例外事項に対応できない問題を「フレーム問題」(frame problem)とよぶが(McCarthy (1968); McCarthy & Hayes (1969); 松原 (1990))、言語の例外事項に対応できない事態は、まさに機械翻訳におけるフレーム問題であった。フレーム問題は、機械翻訳研究を再び停滞させることとなった。

#### ・統計機械翻訳の登場

こうしたなか長尾らの研究(Nagao (1984); Nagao et al. (1985))によって、ルールではなく、対訳データを用いた翻訳が提唱されることとなった(中村 2016, 240)。また1988年にはIBMが文法ではなく、統計によって言語を処理する手法を提唱した。この手法は、それ以前のルールベースではなく、対訳データを統計学的に処理する手法であることから、統計機械翻訳(statistical machine translation)とよばれるようになる。コンピューターの処理速度があがり、デジタルデータを大量に処理できる環境が整ったことも加わり、機械翻訳は大量の対訳データを統計的に処理するモデルにパラダイムシフトしたのである。

#### ・2つの手法の違い

ルールベースと統計機械翻訳の手法の違いは、人間が言語を学習するときの類推で考えてみるとよいだろう。ルールベースによる機械翻訳は、文法や辞書的なデータを人間が人工智能に教えていく手順になることから、人間が第二言語を学習していく過程と似ている。われわれが第二言語を学習する際、とりわけて学校等で習得する場合、文法や単語から学ぶことが多いが、ルールベースはこうした手法と類似している。

これに対して統計機械翻訳は、文法といった事項ではなく、大量の対訳データを統計処理することによってつくられる。こちらのほうは、人間が第一言語を学習するときのアナロジーと親和的である。というのは、ほとんどの場合、人間が第一言語を学ぶのは文法からではなく、周りの人々(親であることが多いだろう)の言語データを吸収することによっているからである。

統計機械翻訳は、データさえあてれば自動的に翻訳システムを生成するので、多言語翻

訳に強いのも特徴である<sup>7</sup>。ルールベースでは、人間が逐一設計しなければならないので、多言語翻訳は大変な労力を必要とする。

#### ・ニューラル機械翻訳

統計機械翻訳の手法は、2000年代まで主流であった。その後統計機械翻訳の手法は、2010年代の半ばごろから機械学習<sup>8</sup>とむすびつくことで、近年ではニューラル機械翻訳(neural machine translation)とよばれるようになってきている。2016年10月には、グーグルが、ニューラル機械翻訳の手法をもちいた新たな機械翻訳『Google ニューラル機械翻訳』(Google's Neural Machine Translation GNMT)を公開している。このニューラル機械翻訳の手法は、従来のものにくらべて大幅に性能があがったといわれている(Wu et al. (2016))。Google 翻訳の精度向上は、各種メディアでもとりあげられた。

ニューラル機械翻訳と統計機械翻訳は、モデルの形成や抱えている課題に違いがあるとされているが(中澤(2017))、パラダイムとしては統計翻訳と変わらない。つまり文法的なルールによって機械翻訳を設計するのではなく、大量の対訳データを処理することで機械翻訳をつくるのである。

現在、機械翻訳は、こうした機械学習がとりいれられたことで、人工知能の最難問とされるフレーム問題の突破が期待されている(松尾(2015))。もしそのような事態になれば、機械翻訳の精度も飛躍的にあがるのが期待できるだろう<sup>9</sup>。

### 3.2 諸言語を媒介する機能はなくせるか？

それでは Ostler がいうように、このまま機械翻訳の技術革新がおこなわれれば、国際共通語のような諸言語の組みあわせを縮減する機能はなくせるのだろうか。筆者には、少なくとも現代の機械翻訳のパラダイムでは、そう簡単にはいかないように思えてならない。それは機械翻訳のしくみからうかがうことができる。

ニューラル機械翻訳は、多言語翻訳の構築に強いという特徴がある。これは言語帝国主義とむきあうには大きな強みである。多言語翻訳の構築が容易であればあるほど、国際共通語

---

<sup>7</sup> ただし、「教師データ・訓練データ」とよばれるデータが同時に必要であり、現在これは人間がつくらなければならない。この点までふくめると、現段階において機械翻訳の構築は、完全に自動的につくられるといえるわけではない。

<sup>8</sup> 人工知能の学習手法のことであり、データをあたえて、繰り返し学習させることで、あたえたデータのなかから特定のパターンをみつけさせることをいう。近年では機械学習のなかでも、深層学習(deep learning)という手法がとりわけて注目されている。

<sup>9</sup> ただし、これには AI 研究者からの懐疑的な意見もある。AI 技術で東京大学入試の合格をめざした研究で知られる新井によれば、現段階での AI 技術による言語処理は、数学という限られた表現手段による制限をうけるため、意味や概念をあつかうことができない。ゆえに現代のパラダイムでは、AI が言語を人間レベルで処理できるようにはならないという(新井(2018))。

のような諸言語を媒介する機能を不要とするからである。

だが問題は、対訳データがなければ、翻訳そのものを構築することがむずかしいことにある。対訳データとは、文字通り「サトウはスズキが好きだ⇔Sato likes Suzuki」といったデータのことだ。こうした対訳データが大量に必要となる。どれぐらい必要になるのかは研究者でも判断しかねているようだが、1000万とも1億とも述べられている（新井 2018, 第2章）。機械翻訳を構築するためには、この規模の対訳データが必要とされているのだ。

対訳データの問題は、大言語同士であれば気にする必要はないのかもしれない。たとえば日本語と英語のような大言語同士であれば、対訳データは大量にあることから、このまま機械翻訳の精度を上げ続けることは可能だろう。しかし、たとえば使用者の数が1万人に満たない少数言語の場合はどうだろうか。このような小規模の言語の場合、政治、経済、社会、文化あらゆる領域において使用されているとは限らず、ほかの大言語で代用されている例が多い。たとえば途上国の多くの大学では、土着語ではなく、多くは旧宗主国の言語が（英語とは限らないが）つかわれている。つまり少数言語は、大言語とくらべて、対訳データが少ない傾向にある。こうした状況となっているのは、ポストコロニアルや言語の近代化などさまざまな要因によって形成されてきた背景があるためだ（こうした諸相の概観については、たとえば Neustupný (2006)）。

たしかに現代の機械翻訳の手法は多言語翻訳を構築しやすい。だが、その機械翻訳の構築に不可欠な対訳データは、すべての言語で均等にあるわけではない。このままでは多言語翻訳は、かなり限定的とならざるをえない。

もっともこの問題は、機械翻訳の研究者も把握している。研究者らが考える解決策は、英語を機械翻訳の仲介言語とすることである。これは、たとえ小規模の言語であっても、英語とだけは対訳データが相対的に多いことがほとんどであることからきている。たとえば、アイルランド語と日本語の対訳データは少ないだろうが、アイルランド語と英語、日本語と英語の対訳データは相対的に多い傾向がある。それならば、アイルランド語と日本語の翻訳も、あいだに英語をはさむことで可能となる。これが解決策として提示される根拠となっている。だが、これはつまるところ、仮に高度な機械翻訳が登場したのだとしても、膨大な言語の組みあわせを縮減する国際共通語の機能が、仲介言語というあり様が変わって残り続けることを意味している。

### 3.3 技術的解決はありうるのか？

技術発展によって、英語を介さない多言語翻訳を実現することは可能ではないかという疑問をもたれるかもしれない。次節へ進む前に少しだけこの点にふれたい。英語を仲介言語にしない多言語翻訳システムは、技術者らも模索している。たとえば Google の研究チームが発表した *zero-shot translation* という発想がある（Johnson et al. (2017)）。これによると人工知能は、機械学習をへていくなかで、独自の内部言語を開発するようになったという。

Interlingua とよばれるその内部言語は、抽象的なイメージに近いと考えられている。

これまでの機械翻訳では、英語が仲介言語としてつかわれていた。だが AI が Interlingua を構成し、仲介言語として使用できるのであれば、対訳データがない言語間での翻訳も可能かもしれない。つまり Interlingua を用いることで、英語を仲介言語としないあり方の可能性がみえてくる。Interlingua を介した多言語翻訳は zero-shot translation とよばれ、Google 翻訳への導入が提唱されている。

もっとも、仮に zero-shot translation のようなモデルが可能であったとしても、それが仲介言語を使用する点は同じである。英語であれ、Interlingua であれ、なんらかの仲介言語の機能が必要とされる。このことは、言語帝国主義を機械翻訳によって解決することを期待する人々が予想していなかった事態をもたらすかもしれない。次節では、そうした点を論じてみたい。

#### 4. 言語帝国主義の潜在化

仮に高度な機械翻訳が登場しても、仲介言語を必要とする。これのなにが問題なのだろうか。もちろん、単に「仲介言語をつかうと、2 言語間の翻訳よりも質が落ちる」ということもできるが、ここで指摘したいのは別のことである。

本節で指摘したいのは、①仲介言語によって生じることと、②機械翻訳の社会的受容の問題である。言語帝国主義は、人々が機械翻訳を使用するレベルでは見えなくなっても、実際の翻訳過程では仲介言語という機能として、残り続ける。こうしたことに人々はより気づきにくくなることが予想される。本稿ではこれを「言語帝国主義の潜在化」として提起してみたい。以下では、まず仲介言語が設けられることによって翻訳にどのような影響を及ぼすのかについて述べる。次に、潜在化とよべる事態を人間社会と技術という観点からみた場合にどのような問題が浮かびあがってくるのかについて論じてみたい。

##### 4.1 言語の視点

仲介言語の問題とは、翻訳される文章に、仲介言語となる言語の視点が相対的に大きな影響力をもつのではないかということである。言語学者の池上嘉彦によると、「世界がどう見えるのか」という視点が、言語もしくは言語使用によって異なってくると述べる（池上(2006)）。本稿では、この世界がどう見えるのかという視点を、言語によって生みだされる、もしくは言語を通じて想起される知覚の慣習としたい。たとえば同じ風景をみても、言語によってその風景を記述する視点が異なるときがありうる。ここで日本語と英語の視点の違いを 1 つみてみよう。川端康成の小説『雪国』の冒頭は、汽車に乗った次の一文からはじま

る<sup>10</sup>。

(1)「国境の長いトンネルを抜けると雪国であった」(川端 1937=1965, 265)。

おそらく、多くの日本語使用者にとってこの一文は、汽車に乗った自分の視点からトンネルを出て雪国がみえはじめる光景を想像するのではないだろうか。日本語の視点では、自分の目線から景色をみるような表現となる傾向がある。ところが英訳された *Snow Country* だと該当箇所は次のようになっている。

(2)The train came out of the long tunnel into the snow country (英訳 1996, 1)。

(2)では、(1)のように自分の視点から雪景色がみえてくる様子を想像することはむしろかしい。むしろ客観的に汽車がトンネルから出てくるような光景を想像するのではないだろうか。

これが誤訳ではないとするなら、なぜ日本語と英語でこのような表現に違いがあるのだろうか。ここには言語が異なると視点が変わることが示唆されている。日本語が自己から客体をみるような傾向であるとするなら、英語は、客体から自己をみる傾向があるという解釈がでてくる。

#### 4.2 サピア=ウォーフ仮説

「言語が異なれば世界が違って見える」という考え方は、現代では Edward Sapir と Benjamin Lee Whorf の「サピア=ウォーフ仮説・言語相対性仮説」(Sapir (1921); Whorf (1956)) が知られている。この仮説は、「どんな言語でも等しく世界を知覚・記述できる」と仮定する生成文法学派と対立関係にある。長らく後者を支持する研究者が多かったが、近年では心理言語学の実験的手法により、サピア=ウォーフ仮説を支持する研究 (Athanasopoulos et al. (2015)) や、生成文法学派の支持者のなかにも言語が異なると視点が変わることを部分的に認めているもの(Deutscher (2010))もある。また認知言語学では、サピア=ウォーフ仮説を穏健化した「弱い仮説」を支持するレイコフのような立場 (Lakoff, (1987)) もある。

この場でどちらが正しいのかを論じることはできないが、少なくとも片方の仮説だけを支持することはむしろかしくなりつつあるということはあるだろう。おそらく諸言語の視点には、共通性もかなりあると思われるが、言語が異なることで世界観が変わることもありうる。雪国と *Snow Country* の冒頭の一文の違いは、日本語と英語の視点が変わることがありうることから生じていると思われるのだ。

#### 4.3 仲介言語によって失われるもの

---

<sup>10</sup> 以下、川端康成の一節とその翻訳の事例は (池上 (2006)) から着想をえた。ただし解釈は、本研究独自のものであるため、責任は筆者におよぶことを明記しておきたい。

『雪国』の一例は、人間の翻訳で生じたことであるが、機械翻訳は人間の対訳データからつくられるため、機械翻訳にも同様のことが起こりうると考えられる。またこれは英語だけに生じることでなく、あらゆる言語間で起こりうる。他方で翻訳によって表現が変わることは、新たな解釈・表現を生み出すことでもあり、ただちにそれが問題といいきることもできない。

だが機械翻訳が、多言語翻訳を実現していくために特定の言語（英語であれ、Interlinguaであれ）を仲介言語にするとなると話は変わってくる。これは翻訳される文章において、仲介言語だけの視点が相対的に強くなっていくと思われるからだ。次の例から考えてみよう。2004年に刊行された『雪国』の新しいドイツ語版 *Schneeland* の冒頭は、以下のようになっている。

### (3) Jenseits des langen Tunnels ershien das Schneeland

（筆者訳：長いトンネルのむこうに雪国があらわれた）

英語版とは異なり、(3)では、日本語の視点を表現しようとしていることがうかがわれる。

注意していただきたいのは、ここで筆者が述べたいのは、ドイツ語のほうが英語よりも日本語の視点をうまく表現できるということではない。また英語で、日本語の視点を表現することは、絶対にできないということでもない。英語でも、ドイツ語のような訳は可能だろう。

ここで問題にしているのは、『雪国』の事例において、もし日本語⇔ドイツ語のあいだに仲介言語として英語がはいっていたら、おそらくドイツ語版は違う訳になっていたのではないと思われることだ。

『雪国』のほんの一例からでもうかがえるのは、仲介言語を介することで、ほかの言語間での翻訳であったら生まれたかもしれないさまざまな表現の可能性を制限してしまうことである。これが仲介言語を設けることによって失われるものといえるだろう。もちろん、こうしたこと自体は、人間の翻訳でもおこってきた。だが高度な機械翻訳が登場するとなると、これまでにない速度で仲介言語を介した翻訳文章がつくられていくことになる。これがどのような効果を、どの程度われわれの言語にあたえるのかは、予想しがたい。

このように述べると、筆者は、「多言語翻訳への要求が理想主義的すぎる」という批判をうけるかもしれない。あるいは「サピア＝ウォーフ仮説を重視しすぎている」と思われるかもしれない。だが、この言語の視点についての議論はこれで終わりではない。言語の視点に関することは、次の機械翻訳がつくられる社会的状況とあわせて1つの問題を形成すると考えられるのだ。

#### 4.4 言語帝国主義の潜在化

仲介言語と言語の視点の議論は、機械翻訳が制作・使用される社会的状況とあわせて、言語帝国主義の潜在化とでもよべる事態を形成すると予想される。これを人間社会と技術をめぐる問題としてとらえてもよいだろう。この問題を論じるためには、Lessig の議論か

らはじめる必要がある。

Lessig は、インターネットを事例にアーキテクチャという問題を批判的に考察している（以下は、Lessig (2000)によっている）。彼が述べるアーキテクチャとは、人々のふるまいを決める設計である。たとえばアメリカの病院や劇場には、携帯電話の電波を遮断するようにつくられている箇所があるという。これはアーキテクチャによって、人々に特定の行動をさせないように（この場合は、携帯電話がかけられないように）規制されていることを意味している。

アーキテクチャによる規制のなにが問題なのか。それは一般の人々は、どのようなアーキテクチャが設けられているのかを十分に知る機会をあたえられず、アーキテクチャの設計者だけが規制を構築することができることだ。たとえ問題のあるアーキテクチャの規制が設けられていたのだとしても、一般の人々がアーキテクチャを設計しなおすことはできず、なにより自分たちのふるまいがどう規制されているのかに気づかないことがほとんどである。Lessig が問題にしたのはインターネットであるが、この論理は、機械翻訳と、その使用者との関係にみいだすことも可能だろう。

機械翻訳をつくるのは技術者であり、一般の人々は事実上その制作に関わることはできない。どの言語間で仲介言語が使われているのかも、われわれが知るのは容易ではない（こうした情報は、公開されていないことが多い）。

「言語が異なると世界が違ってみえる」というサピア=ウォーフ仮説をどこまで支持できるのが未確定なのだとしても、われわれがこうしたことを認識し、判断することなく、技術者たちによって人々の言語使用のふるまいが決められてしまっている。そして多くの人々は、このことに気づくこともないまま、仲介言語の視点の影響を受けることになるかもしれない。こうしたことはこれまでの言語帝国主義とは異なる点である。

すでに多くの人々が使用する『Google 翻訳』等では、英語を仲介言語とする翻訳の組みあわせは存在している。言語帝国主義の潜在化は、すでにはじまっているといえるだろう。そしてこれは *Interlingua* のような人工知能の内部言語が仲介言語となっても、問題の構造自体は変わらない。というのは *Interlingua* が抽象的なイメージであっても、その言語には、なんらかの視点があると思われるからだ。おそらくその視点は、*Interlingua* を構成していくために用いられる言語データの影響を強く受けたものではないだろうか。そうであるならば、それはあらゆる言語にとって中立といえるものではないことになり、英語を仲介言語とするのと同様の問題を抱えることに違いはないだろう。

#### 4.5 社会と技術

つまるところ人々がどのような機械翻訳を望むのかという議論がないまま、設計者の判断によって仲介言語が決められている。高度な機械翻訳ができれば、翻訳された文章は、飛躍的に増加し続けるだろう。だが、知らぬまに翻訳される文章のなかで、仲介言語の視点が

強く残るかもしれない。そしてこれは、人々が機械翻訳を使用するユーザーレベルでは、気づきにくいことでもある。本稿で指摘したいのは、高度な機械翻訳が登場するのだとしたら、こうした事態が予想されるのではないかということだ。

機械翻訳という技術が人間社会を変えるのか。たしかに表層的には、機械翻訳により、国際共通語（英語）の覇権的地位は、低下ないしは消滅したようにみえるのかもしれない。だが、このことをもって、言語帝国主義は終わったということとはできない。言語帝国主義は、水面下における仲介言語の視点の拡大となって残り続けるだろう。機械翻訳が言語帝国主義を終わらせるのかという問いに対する答えは、否である。

Ostler を含めて、言語帝国主義とむきあう議論のなかで機械翻訳に期待する人々が、こうした事態を望んでいるとは考えにくい。そうであるならば、このような問題に対して、どのような対応がありうるのかということこそ、われわれが考えていかねばならない課題ということになる。

## 5. おわりに

本稿では言語帝国主義とされる状況を、機械翻訳が終わらせることができるのかという問いから論じてきた。結論は否である。むしろ言語帝国主義は潜在化することで、自らが特定の言語の影響下にあることを多くの人々が認識しにくい状況になることが予想される。

このような状況に対して、さまざまな対応を考えていく必要があるだろう。機械翻訳に仲介言語が不可欠であるのならば、それはどの言語であるべきか。どの程度の情報公開が必要になるのか。機械翻訳を使用すべきではない局面があるとすれば、それはどういったときだろうか。グーグルのようなグローバル企業は、機械翻訳をつくることに大きな影響力をもっているが、われわれは、いかにそれを統制することが可能なのか。現状は、こうした議論がまったくといっていいほど存在しないまま、技術発展だけがすすめられている。

また機械翻訳と言語帝国主義というテーマについて、本稿で論じているのはごく一部であることも明記しておきたい。たとえば本稿では英語にばかり焦点をあててきたが、機械翻訳は英語以外の大言語の力を増す作用もあると思われる。現代でも大言語さえ学べば、さまざまな情報にアクセスしやすくなる状況があるが、機械翻訳はそうした傾向を強められると思われる。これは少数言語使用者にとって、自身らの言語を英語や英語以外の大言語に切りかえる強い動機になると考えられる。そのため、いわゆる少数言語が大言語に吸収されていく「言語ジェノサイド」問題（Skutnab-Kangas (1999)）を加速させるかもしれない。

たしかに機械翻訳技術の進歩は著しい。けれどもそれにともなって、これまで考えられてこなかった問題が、次々と噴出することが予想される。今後機械翻訳のユーザーとなるわたしたちは、ただ新たな技術を享受するばかりではなく、このような技術とどうつきあっていくのかを考えていく構えをもつことが必要なのだ。

## 【文献リスト】

・日本語

- 新井紀子. 2018. 『AI vs. 教科書が読めない子どもたち』 東洋経済新報社.
- 池上嘉彦. 2006. 『英語の感覚・日本語の感覚』 NHK ブックス.
- 岡崎晴輝、鎌田厚志. 2016. 「エスペラント語の政治理論」『政治研究』(63)、九州大学政治研究会.
- カルヴェ, L.J. (著)、砂野幸稔、今井勉、西山教行、佐野直子、中力えり (訳) . 2010. 『言語戦争と言語政策』 三元社.
- 川端康成「雪国」(『川端康成集 30』筑摩書房) . : Kawabata, Yasunari, translated by Edward G. Seidensticker, (1996), *Snow Country*, New York: Vintage Books. : Yasunari Kawabata , Aus dem Japanischen von Tobias Cheung, (2004), *Schneeland*, verlag: Bibliothek Suhrkamp.
- 木村護郎クリストフ. 2013. 「英語のメガホンをとれ!: 世界の英語化による公正のすすめ」『社会言語学』(13)、社会言語学刊行会 : 187-193.
- 木村護郎クリストフ(編著). 2016. 『節英のすすめ: 脱英語依存こそ国際化・グローバル化対応のカギ』 萬書房.
- 小林雅一. 2015. 『AI の衝撃 人工知能は人類の敵か』 講談社.
- サンダルソラ, トール (著)、塚本茂蔵・吉田卓 (訳) . 2010. 『言語帝国主義: 英語と世界制覇の夢と現実』 幻冬舎.
- 中澤敏明. 2017. 「機械翻訳の新しいパラダイム: ニューラル機械翻訳の原理」『情報管理』60(5)、国立研究開発法人科学技術振興機構 : 299-306.
- 中村哲. 2016. 「多言語音声翻訳技術の最前線」砂岡和子、室井禎之『日本初多言語国際情報発信の現状と課題: ヒューマンリソースとグローバルコミュニケーションのゆくえ』朝日出版社 : 233-250.
- 長尾真. 1986. 『機械翻訳はどこまで可能か』 岩波書店.
- 長尾真. 2016. 「巻頭言 言語の壁を乗り越え相互理解を深めるために」砂岡和子・室井禎之 (編著) . 『日本発多言語国際情報発信の現状と課題: ヒューマンリソースとグローバルコミュニケーションのゆくえ』朝日出版社 : 1-4.
- 新田義彦. 2012. 『機械翻訳の原理と活用法——古典的機械翻訳再評価の試み』 明石書店.
- 松尾豊. 2015. 『人工知能は人間を超えるか ディープラーニングの先にあるもの』角川選書.
- 松原仁. 1990. 「一般化フレーム問題の提唱」マッカーシー, J.、ヘイズ, P.J.、松原仁 (1968-9=1990) 松原仁、三浦謙訳『人工知能になぜ哲学が必要か: フレーム問題の発端と展開』 哲学書房.

- 三浦信孝、糟谷啓介（編）. 2000. 『言語帝国主義とは何か』 藤原書店.
- 吉川寛. 2016. 「国際英語論とは」 塩澤正、吉川寛、倉橋洋子、小宮富子、下内充『「国際英語論」で変わる日本の英語教育』くろしお出版.
- 渡辺太郎、今村賢治、賀沢秀人、Neubig, Graham、中澤敏明（共著）、奥村学（監修）. 2014. 『機械翻訳』 コロナ社.

・ English

- Athanasopoulos, P., Bylund, E., MonteroMelis, G., Damjanovic, L., Schartner A., Kibbe, A., Riches N., and Thierry, G. 2015. "Two Languages, Two Minds: Flexible Cognitive Processing Driven by Language of Operation, *Psychological Science*, Vol 26, Issue 4:1-17.
- Crystal, D. 1997=2003. *English as a Global Language*, Cambridge: Cambridge University Press.
- Deutscher, G. 2010. *Through the Language Glass: Why the World Looks Different in Other Languages*, New York: Metropolitan Books. (=ガイ・ドイッチャー著、椋田直子訳, 2012. 『言語が違えば、世界も違って見えるわけ』 合同出版).
- Fiedler, S. 2010. Approaches to Fair Linguistic Communication, *European Journal of Language Policy*, 2(1):1-22.
- Kachru, B. 1976. Models of English for the Third World, *TESOL Quarterly*, 10(2), 221-239.
- Jenkins, J. 2007. *English as a lingua franca: Attitude and identity*, Oxford: Oxford University press.
- Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z. and Thorat, N. 2017. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. <https://arxiv.org/abs/1611.04558> (Accessed May 11, 2018)
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things*, University of Chicago Press: Chicago. (=ジョージ・レイコフ著、池上嘉彦・河上誓作訳, 2016. 『認知意味論: 言語から見た人間の心』 紀伊国屋書店).
- Lessig, L. 2000. *Code: And Other Laws of Cyberspace*, New York: Basic Books. (=ローレンス・レッシング著、山形浩生・柏木亮二訳, 2001. 『CODE—インターネットの合法・違法・プライバシー』 翔泳社).
- McCarthy, J. 1968. Programs With Common Sense, In *Semantic Information Processing* eds. M. Minsky, 403-418. Cambridge: MIT Press.
- McCarthy, J. & Hayes, P. J. 1969. Some Philosophical Problems From the Standpoint of Artificial Intelligence, *Machine Intelligence*, 4:1-51.

- Nagao, M. 1984. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. *Artificial and Human Intelligence*, 173-180.
- Nagao, M., Tsujii, J. and Nakamura, J. 1985. The Japanese Government Project for Machine Translation, *Computational Linguistics*, 11, 2-3:91-110.
- Neustupný, J.V. 2006. Sociolinguistic Aspects of Social Modernization, In *Sociolinguistics An International Handbook of the Science of Language and Society*, eds. U. Ammon, K. Mattheier, and P. Trudgill, 2209-2223. Berlin: Walter de Gruyter.
- Ostler, N. 2010. *The Last Lingua Franca: English Until the Return of Babel*, United Kingdom: Walker & Company.
- Phillipson, R. 1992. *Linguistic Imperialism*, Oxford: Oxford University Press.
- Sapir, E. 1921. *Language: An Introduction to the Study of Speech*, United States: Harcourt, Brace.
- Schutter, H. De, Robichaud, D., eds. 2016. *Linguistic Justice: Van Parijs and His Critics*, Abingdon-on-Thames: Routledge.
- Seidlhofer, B. 2001. *Understanding English as a Lingua Franca*, Oxford: Oxford University.
- Skutnab-Kangas, T. 1999. *Linguistic Genocide in Education - or Worldwide Diversity and Human Rights?*, Mahwah: Lawrence Erlbaum Associates.
- Smith, L.E. ed. 1983. *Reading in English as an International Language*, Oxford: Pergamon Press.
- Van Parijs, P. 2012. *Linguistic Justice for Europe and for the World*, Oxford: Oxford University Press.
- Whorf, B. 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, Cambridge: MIT Press.
- Wu, Y. et al. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. <https://arxiv.org/abs/1609.08144> (Accessed May 11. 2018).

## **Can The Machine Translation end Linguistic Imperialism?: Thinking from its Mechanism**

NISHIJIMA, Yu

Postdoctoral Fellow, Graduate School of Global Studies, Sophia University

This paper considers whether or not advanced machine translation can end the English

unipolar situation from its mechanism, assuming it will be realized in the future. Machine translation is expected to end the unipolar situation of English but this article concludes it cannot happen. The modern paradigm of machine translation is called “neural machine translation” and is made through learning human language data. This will lower the influence of English if translation is between large languages with large amounts of data. However, in the case of minor languages with poor data, the influence of English is still expected to remain.