

# 言語処理で研究すべきことはまだまだ多い

There is still much work to be done in the natural language processing field.



京都大学名誉教授  
**長尾 真**

1997～2003年京都大学総長、2007～2012年国立国会図書館長、2005年日本国際賞、レジオンドヌール勲章、2008年文化功労者、平成30年文化勲章受章

Japioにおける産業日本語研究会が始まって今年は19年を経過した。その間、機械翻訳における誤りの少ない日本語表現はどのようなものであるかについて、種々の観点から研究がなされてきた。また機械翻訳方式も文法中心の翻訳方式から用例に依存した用例翻訳、これに統計的観点を導入した統計翻訳方式を経て、数年前から大量の対訳データをもとに深層学習機能を入れたニューラル翻訳方式へと発展し、翻訳能力が抜本的に良くなってきている。

このニューラル翻訳方式を改良してゆけばまだまだ機械翻訳の質が良くなって行く余地があるだろうが、それだけでは解決できない課題も依然として多く存在する。その代表例は、指示詞が何を指しているか、埋め込み文における主語は通常省略されるがそれを正しく復元することができるか、といった文脈を正しく読む能力に欠けるという問題がある。これは用例をさらに増やすことである程度解消できるだろうが、根本的にはある種のアルゴリズムを導入して推定をする必要があるだろう。一つの文が極端に長い場合、その中に存在する複数の述語がどの範囲を支配しているかを決定することは非常に難しく、これを用例の参照で解決しようとするには無理がある。人間は文脈的なある種の論理で解決していることは間違いなく、そのアルゴリズムを明確化して、ニューラル翻訳を行う前に適用して文の構造を確定しておくことが求められる。

知識を用いて推論することが求められる場合もある。今年の春の日本経済新聞にあった例が面白かったが、正確な月日と書いた人の名前（アメリカ人）をメモしてお

かなかったのが悔いられるが、要するに次のようなものだった（文例は少し違っているが）。

荷物を持っている人に「鞆に入れたらどうですか」と言ったとき、「これが大きすぎるのです」という返事だったら、「これ」は荷物であり、「これが小さすぎるのです」と言ったら、「これ」は鞆を指すという判断が必要である。

「誤訳の案内板、訪日客が困惑」という新聞記事が出たことがある（日本経済新聞、2019年5月11日夕刊）。例えばビルの出口の案内で「お帰り口」というのが中国語では「あなたが出て行け」という表現に機械翻訳され掲示されていたりする。機械翻訳の結果は母国語を話す人にチェックしてもらおうか、少なくとも一度逆翻訳（翻訳されたものをもう一度元の言語に機械翻訳して戻してみる）を試みて、意味の通らないものになっているかどうかを調べることはやってみるべきだろう。

このようなことを考えると、同じことが異なった表現で表されている場合、それが内容的に同じであることをどうして知ることができるかという言語学的に興味のある課題に直面する。「出口」と「お帰り口」だけでなく、「全席禁煙」と「喫煙席はありません」はまだ何とかなるだろうが、「それはお塩ですか」と「そのお塩を取ってください」とが同じことを意味するとする語用論に関する問題を機械翻訳に取り入れるところまで考えねばならないだろう。

言葉は常に変化しているし、新しい言葉がどんどん出てきている。若い人たちでは常識なのだろうが中年以降

の人達には理解できないような短縮語、外来語、などが新聞や雑誌にあふれている。学術用語、産業用語、スポーツ用語、その他それぞれの専門分野の新しい用語が常に出てきているが、これらに対しては国立国語研究所などの中立機関で常に集めて一般に対して提供するといった活動が必要である。言葉は文学のためだけのものではなく、社会のあらゆる人にとってのものであることを改めて認識することが大切であろう。

最近ではワープロも簡単な学習機能を持ち、良くなってきているが、産業日本語あるいは機械翻訳のためのワープロといった観点でもう少し工夫されたものを開発してはどうだろうか。例えば一つの文がある程度以上の長さになると警告を出すとか、文の係り受け構造に曖昧さがあるから語順を入れ替えてはどうかとか、並列句の範囲のチェックといったことの確認、そして1文を書き終わった段階で省略された主語や目的語を推定したり、代名詞の指すものを挿入して示すことによって正しい訳文が出ることを保証したりするソフトウェアを入れたワープロを作るといった工夫である。このようなワープロからの警告を受けて文章を改良するのはあくまでも書き手に任せるのである。このようなシステムは値段が高くなり、使う人にも負担をかけることになるが、機械翻訳の結果を時間をかけて後修正するためのコストと比較すれば、決して否定すべきものではないだろう。このような観点からすれば、ほぼ完成してもう何もしなくてもよいと思われていたワープロにも改良の余地はいろいろあるのではないだろうか。

曖昧な文章は特許文章などでは特に問題となるだろう。意図的にあいまいな表現にすることは許されない。そういったところから、日本語文章をできるだけ曖昧で

ない形に書くことが求められ、我々の研究会でもいろいろと研究されてきた。全く曖昧さを含まない文章といったものは作れないが、かなりの程度曖昧さを排除した文章を作るための指針をこの研究会で作れたことは大きな業績と考えられる。これからはこの指針をワープロに入れ、特許文章を作る人達が実践するよう普及する努力のフェーズになってきている。

2020年東京オリンピック・パラリンピックでは10前後の言語の音声通訳器が導入され、外国から来る人たちが自由に使えるようにすると言われている。これを使えば会場への道を迷ったり急に病院に駆け込みたいといったときにも力を発揮し、安心してスポーツを楽しむ、観光旅行ができるだろう。その場合の多くは質問応答のモードである。しかし例えば「遠くでなくて今から行って間に合う半日くらいで楽しめる競技が良いのだが」といった質問の時は、「何がお好きですか、室内競技がいいのか屋外のほうが良いのかどちらですか」といった確かめのための逆質問になるだろうから、単純なQ&Aシステムでは済まなくなる。したがって種々の知識を用意しておかねばならないし、相手が大人なのか子どもなのか、また男女の区別などなど、相手の認識も必要になってくるだろう。

将来地球上のあらゆる人が何十言語も扱える音声通訳器を持ち、いろんな言語の人達と気楽に対話できるようになれば、世界中の人達の相互理解が進み、深刻な対立から忌まわしい戦争になるということが無くなってゆくだろう。人生のほとんどを言語処理、機械翻訳の分野にささげてきた人間にとっては、そのような理想的な世界が実現することを念願せざるを得ないのである。

Japio YEAR BOOK 2019 には、昨年10月26日に平成30(2018)年度文化勲章を受章されました、京都大学名誉教授の長尾真先生から特別寄稿をいただけることとなりました。

長尾真先生には、日本特許情報機構が主催する産業日本語研究会に、その発足当初から現在まで顧問としてご指導いただいております。このたびは、先生のご受賞を心よりお祝い申し上げますと共に、当機構における産業日本語、機械翻訳等研究への永年にわたる多大なる先生のご支援に対し、厚く御礼申し上げます。

日本特許情報機構(Japio)一同

#### <長尾真先生ご略歴>

長尾真先生は、京都大学教授、京都大学総長を務められ、ご退官後も、独立行政法人情報通信研究機構理事長、国立国会図書館長等を歴任されました。この間、画像および言語などの情報メディアを用いた知的情報処理に関する最先端の研究を行われ、画像処理、自然言語処理、機械翻訳、電子図書館等の分野において優れた実績を挙げられました。

また、言語処理学会初代会長、機械翻訳国際連盟初代会長など多数の学会における要職を歴任されるとともに、これまでの一連の研究業績により、平成5年にIEEE Emanuel R. Piore 賞、平成9年に紫綬褒章、平成15年にACL Lifetime Achievement Award、平成17年に日本国際賞、平成30年には文化勲章を受章されています。