グローバル時代の情報収集と情報発信

Information distribution and assimilation in the global era

東京大学名誉教授／マイクロソフトリサーチアジア研究所首席研究員 **辻井　潤一**

PROFILE 国際機械翻訳協会（IAMT）およびアジア太平洋機械翻訳協会（AAMT）元会長、AAMT／Japio 特許翻訳研究会委員長、国際計算言語学会（ACL）元会長、国際計算言語学委員会（ICCL）永久メンバー

jtsujii@microsoft.com　TEL +86-13520978845

1 はじめに

昨年まで、Japio YEAR BOOK には私が専門とする自然言語処理や機械翻訳の技術が特許情報処理の分野でどのように活用できるかに焦点を当てて議論をしてきた。この方向での議論は、ここ数年の私の寄稿論文を参照してもらうとして、本年年度は、国外から日本を見るという立場から、言語処理、機械翻訳といった技術が日本の置かれた研究、技術開発の環境にどのように貢献し得るかについて、整理してみようと思う。

もちろん、日本の科学技術が国際的にどのような地位を占めているかを系統的に調べて議論するわけではないので、私個人の経験からの、きわめて偏った見方からの議論になっていること、あらかじめ了解していただきたい。

2 グローバル化

経済、社会活動の幅広い範囲でグローバル化が急速に進展していることは、あらためて指摘するまでもない。このグローバル化は、研究や技術開発の分野でも大きな影響を与えている。実際、研究者同士の国際的な共同で交流、国境を越えた技術開発は、日常茶飯事になっているといってよい。

私の最近の発表論文でも、共著者が日本人のみという論文はきわめて稀なものになっている。専門（言語処理、機械翻訳）の関係上、私は、平均の技術者や研究者よりも、言葉の問題には敏感になっているが、日本語で書いた学術論文は、ここ数年皆無である。日本語で書かれた論文は、日本以外の国の研究者にとっては、日本が非常に先行しているごく少数の分野を除いては、無に等しいものになりつつある。日本語で書かれた論文が、世界の科学技術研究にインパクトを与えることはほとんどない。

重要なことは、この英語を中心にした情報流通は、日本の情報発信力に影響を与えているだけではない。日本の情報収集力にも負の影響を与えており、日本の研究開発力そのものに非常に大きな負の影響を与えている。

3 グローバル化と情報収集力

インターネットが普及する以前、たとえば、20 年前と比べると、情報の流通速度が飛躍的に増大している。世界のいろいろな場所で開催されている国際学会の論文集は、学会に参加していなくても、インターネットを通じて、即日にダウンロードして読むことができる。

学術論文誌も、購読料で論文誌を運営するビジネスモデルから、論文を投稿する研究者が費用を負担し、論文購読者は、無料で論文をダウンロードできるオープンアクセスのビジネスモデルへと移行しつつある。オープンアクセスの論文は、だれでも自由にダウンロードできて自由に読むことができる。オープンアクセスのビジネスモデルには、賛否両論がある。ただ、研究成果を出来るだけ多くの人に迅速に伝達したい、世界の研究者にインパクトを与えたいという研究者の本能的な願望を考えると、オープンアクセス化の流れは、今後、ますます加速

する。

論文誌のオープンアクセス化と同様な流れに、ソフトウェアのオープンライセンス化もある。論文誌のオープンアクセス以上に賛否があるが、この流れも、できるだけ多くの人に自分のソフトウェアを使ってもらいたいという開発者の本能的な願望、また、集団的なソフトウェアの開発とメンテナンスの必要性を考えると、やはり不可避的な流れであろう。また、ソースコードのオープンライセンス化までは行かずとも、実行コードのダウンロードを自由に許す動きは、さらに広範に見られる。

手短にいうと、研究や技術開発のための情報やツールがインターネットに氾濫しており、研究組織や研究者の物理的な所在位置とは無関係に、これらの情報やツールにアクセスできる。いわば、物理的にどこに住んでいようと、あるいは、研究者の所属組織とは無関係に研究開発を行うことが可能になってきている。

インターネット環境が存在しなかった、たとえば、20 年前には、物理的にどこにいるか、どのような研究組織（大学、公的研究所、大企業の研究所）に所属しているかが研究環境を大きく左右していた。米国、ヨーロッパや日本という、研究開発に先行している地域の主要な研究組織に所属していることが、優れた研究成果を挙げる必須の条件となっていた。これに対して、インターネット時代は、このような条件の持つ意味が大きく低下してきている。物理的な位置、所属組織の重要性の相対的な低下に伴って、研究者個人の情報収集力や情報発信力、研究推進能力といったものの重要性が強くなってきている。

もちろん、インターネットに存在する情報が、研究基盤のすべてではない。研究開発の競争力を高めるには、研究情報以外の基盤（たとえば、実験装置）構築のための適切な政策と投資は不可欠であろう。また、装置など物理的な拘束がある以上、研究者の能力が完全に物理的な所在地や所属組織から自由なわけではない。

インターネットを使った会議が普遍化しても、研究者同士が日常的に接触し Face-to-Face の会話が新たなアイデアを生み出す原動力になることは、シリコンバレーなどの例から明らかであろう。物理的な位置の重要性は、もちろん、なくなったわけではない。

ただ、ここで指摘しておきたいのは、物理的な所在地の意味が相対的に低下し、インターネットを通じた研究情報やツールの収集能力、物理的に散在している研究者とのコミュニケーション能力といった個人の能力の重要性が飛躍的に増大したことである。シリコンバレーの例でも、所属機関を超えた研究者個人の交流が新しいアイデアの創生に果たす重要性が指摘されている。

このような個人の情報収集能力やコミュニケーション能力に、英語の能力が深く関与していることは明らかであろう。国際言語としての英語の能力は、情報発信力だけでなく、日本の研究者、技術者が研究を進めていく上での情報収集力に大きな影響をもつ。

4 英語能力と情報収集力

さて、日本の研究者、技術者の英語能力である。私は、これまで京都大学、東京大学という、いわばエリート集団に属する学生と研究を進めてきた。彼らの英語能力は、日本の学生の平均よりもはるかに高い。また、私の学生時代に比べると、彼らの英語能力の高さに感心することも多い。

しかし、そのような彼らでも、ヨーロッパや他のアジア諸国の学生の英語力に比べると、総体としては、正直、見劣りする。訓練の問題であろうが、大学院の修士課程に入学した段階で彼らに英語の論文を読ませると、一つの論文を読むのにかなりの時間がかかる。数日かかるのも、稀ではない。

このことは、情報収集力、最新の論文や研究の道具としてのソフトウェアへのアクセス能力の点で、非常に損をしている。しかも、科学技術が成熟した日本では、日本語のよく書かれた教科書や解説、日本語の学会誌などが存在し、日本語だけでそれなりの情報アクセスが可能なために、修士以降の研究においても、英語での情報収集能力がそれほど伸びない。研究者として成長していく過程での情報収集能力の訓練が不足する。国内での研究者人口が大きく、それなりの水準を持つ日本の環境が、情報収集や情報発信という、研究能力の重要な要素の訓練には負の影響を及ぼす。

このような環境にない国では、大学以前の教育や大学院以降の研究における英語による情報アクセスが必須な

ために、研究者は、英語での情報アクセス能力を若いときから自然に身に付けていく。この能力の差が、研究や技術開発、あるいは、国際的なマーケティングや技術動向の把握における物理的な位置、国境の意味が希薄化している現在において、日本が非常に不利な状況に置かれている原因の一つではないかと思われる。

5 日本の状況

たとえば、私が勤務した英国の大学（マンチェスター大学）では、博士課程の学生であれば、自分の研究に関係する論文 10 篇や 20 篇は、1 週間程度で読みこなす。いわゆる流し読みで、自分の研究に関連した論文が非常に迅速に読める。この能力は必ずしも英語が母国語の学生だけではなく、博士を取得する前提条件のようなものとなっている。

日本人の学生でも、自分の研究に関連した日本語の論文であれば、その要点だけを読んでいくことで、20 篇程度の論文を読みこなすのはそれほどの苦労はない。ただ、これが英語論文ということになると、読み込む速度が格段に低下する。研究や技術開発のスピードが加速している現在、最新の英語論文の読解速度の遅さは、日本での研究、開発に明らかに不利になっている。英語と日本語でアクセスできる情報は、量と質の両面で非常に大きな差がある。

英語能力の問題は、かならずしも論文からの情報収集にとどまらない。技術動向の把握に不可欠な特許情報はもちろん、前述のオープンライセンスのソフトウェアやダウンロード可能なソフトウェアの説明やマニュアル類も、その多くは英語で書かれている。このために、有効なソフトウェア類が氾濫しているにもかかわらず、日本のソフトウェア開発者が利用していない可能性も高い。ソフトウェアを共有するためのプラットフォーム作りが世界的に広がっている中で、このことが日本のソフトウェア産業に負の影響は非常に大きい。

世界的に使われている Ruby の開発者である松本行弘さんの「Ruby を広げる上で英語でのマニュアル作りが非常に重要な役割を果たした」というコメントは、世界標準を作っていくための情報発信での英語の重要性を示している。この情報発信力と同様に、あるいは、それ以上に、情報アクセスでの英語能力の重要性がグローバル時代の研究者には要求されるようになっている。英語が標準のコミュニケーション言語となっているインドが、高いソフトウェアの生産性をもち、この産業分野でリードする位置にあることは、ソフトウェア開発者の英語能力の重要性を示すものであろう。

6 大学の国際化

この原稿を準備していたときに、英国の Times Higher Education で世界の大学ランキングが発表された。日本でも報道されていたので、その記事を読まれた方も多いであろう。

その記事によると、「東大が 23 位でアジアでは 1 位の座を守った。順位は下げたが、京大（59）、東工大（141）、阪大（157）、東北大（165）と続いている」とのことであった。Times のウェッブサイトで調べてみると、確かにこの記事のとおりであるが、他のアジア諸国の大学も、シンガポール国立大学（25）、香港大学（43）、北京大学（48）、精華大学（49）、ソウル大（50）、香港科学技術大（51）、KAIST（韓国、52）、南山科技大学（シンガポール、61）、Postec（韓国、66）、香港中文大（129）といった具合に高いランクに入ってきている。日本で 2 位の京大（59）よりも上にすでに 7 つ、アジア諸国の大学がランク入りしている。これらのアジア諸国の大学は、年々ランクを上げていることから、日本の大学のアジアでの地位は、このままではさらに低下していく。

さらに深刻なのは、Times が発表している技術系大学のランキングである。このランキングでは、東大は世界で 27 位にランクされ、シンガポール国立大（13）、香港科学技術大（21）、精華大（24）、KAIST（韓国、25）と 4 つの大学の後塵をはいし、ソウル大（29）、南山科技大（シンガポール、33）が東大に肉薄している。

新聞のコメントは、ここ数年の決まり文句で、「日本の大学は留学生、外国人教員、海外との共同研究などで悪い評価をされており、大学の国際化が必要」とある。

この新聞のコメントは間違ってはいない。ただ、この

問題は、留学生や外国人教員を増やしたり、国際共同研究のための研究資金を増やしたり、といった提言がされる。もちろん、これらの施策は必要であるが、このような表層的な、いわば、お化粧的な施策だけでは、日本や日本の大学が抱える問題を解決できない。

研究や技術開発がグローバル化するなかで、日本の研究教育の環境が従来のローカルな状態とどまっていることの解消が、より重要になっている。日本が技術先進国であり、したがって、その日本の研究組織に所属しているだけで、情報収集力や情報発信力で優位に立てた時代は、すでに終焉している。物理的な位置とは無関係に最先端の情報に自由にアクセスでき、個人の研究者が自由に国境を越えて組織を移動できる現在では、所属機関が持つ情報アクセスの優位性は、著しく低下している。重要なことは、研究者個人の情報収集能力、情報発信能力を向上させ、その結果として、国際的な共同研究が自然と行われていく、あるいは、日本の組織に海外からの研究者、技術者が流入するという環境を構築することである。

7 言語処理、機械翻訳の必要性

さて、日本がおかれた、このような不利な状況に、改善する方策は、あるのだろうか？　特に、言語処理や機械翻訳という技術は、どのように貢献できるのだろうか？

短絡的な答えは、機械翻訳の技術が進展することで、英語の情報世界と日本語の情報世界の格差は徐々に解消し、日本の研究者や技術者の情報収集、情報発信の能力の問題はなくなる、ということであろう。実際、このユートピア的な議論も、正しい側面を持つ。現在の不完全な機械翻訳システムでも、一部の技術者は、たとえば、オープンソースやウェッブサービスの英語で書かれたマニュアルを読むのに活用している。研究者、技術者の情報収集力を向上させるのに貢献している。特許や論文の機械翻訳が、技術動向の把握に貢献し始めていることも確かである。

ただ、現在の機械翻訳システムをこの目的に使った経験からは、テキストの流し読みには有効であるが、本当の理解には英語の原文を読む必要がある。この原文を理解するためのツールは、現在のところ決定的に欠如している。機械翻訳と英語読解支援システムの有機的な融合は不可欠であろう。

上の観察は、情報発信での機械翻訳の使用にも当てはまる。現在の機械翻訳は、日本語から粗い下訳を作るのには使えるが、下訳から情報発信に使える英語テキストへの変換は容易ではない。現在のところ、翻訳専門家の関与が不可欠となっている。研究者、技術者が直接英語テキストを作る英文作成支援システムと、現在の機械翻訳技術の融合が不可欠であろう。

さらに重要なことは、このような英文読解支援、英文作成支援システムの有効性は、それを使う技術者、研究者の英語能力そのものに大きく依存する。結局のところ、技術者、研究者個人の英語能力を向上させていくための訓練をどうするかが最大の課題となる。とくに、それなりに大きな研究者、技術者人口を抱える日本は、日本語だけである程度の情報収集と情報発信ができてきたことで、この訓練の過程が決定的に不足している。

この問題は、日本という社会システムの中での教育という、技術を超えた領域に入り、技術以前の問題が大きい。ただ、言語処理の技術を言語教育の中に取り込んでいこうという研究が活発化しつつある。ここでは、大学以降の研究者、技術者の訓練課程にいかにこのような技術を導入し、英語と日本語の世界を自由に移動できる技術者や研究者と養成すること、また、そのための技術を開発していくことが重要となる。