

産業日本語プラットフォーム委員会の活動を始めるにあたって

豊橋技術科学大学情報メディア基盤センター教授 井佐原 均

PROFILE

電子技術総合研究所、情報通信研究機構（元、通信総合研究所）を経て、現職。産業日本語研究会世話人代表。産業日本語プラットフォーム委員会委員長。アジア太平洋機械翻訳協会会長。機械翻訳国際連盟会長。



1 はじめに

平成22年2月に第1回産業日本語研究会・シンポジウムが開催され、産業日本語というキーワードの下で、情報技術を用いて、我が国の産業競争力の強化を目指す動きが始まった。本稿では、情報技術と産業競争力とのかかわりについて概観した後、国際競争力の観点から特に重要となる機械翻訳の現状について述べ、我が国の現状を踏まえて、産業日本語研究会の実働部隊となる産業日本語プラットフォーム委員会の方向性を紹介する。

2 情報技術と産業競争力

情報処理技術の発展は、その技術を製品化する情報産業自体の競争力の強化に直結するとともに、情報処理技術をツールとして用いることにより、より広い産業における競争力の強化に資すると言われている。これは言語を対象とする情報処理技術（自然言語処理技術）においては、より顕著であり、たとえば機械翻訳技術の発展は、機械翻訳システム単体としての競争力を高めると共に、他の産業におけるマニュアルや広報文書の翻訳の高精度化と効率化が実現されることにより、その産業の国際競争力の強化に寄与できる。

残念ながら、現在の自然言語処理技術は完璧なものではなく、自然言語処理技術が我が国の産業全体の国際

競争力の強化に役立つ部分は限定的である。このような制約の下で、我が国の産業の国際競争力の強化の観点から、情報処理技術がどのように寄与できるかをきちんと検討・評価すべきであろう。

我が国の国際競争力の強化と日本語文書の計算機処理の能力向上を考える場合、いくつかのアプローチが考えられる。ひとつは現在の自然言語処理技術は自由に記述された日本語文書を解析するには不十分であるため、日本語を何らかの形で制約し、文書解析の精度が上がるような、翻訳しやすい、あるいは検索しやすい文書を作成しようというアプローチである。いわゆる制限日本語の考えであり、産業日本語のひとつの方向であろう。メタデータの付与による検索や情報獲得の高精度化も考えられる。別のアプローチとしては、最初に記述される日本語文書のテキストには制約をかけず、その文書を変換して、機械翻訳システムにかかりやすい日本語文書を作成するという考えもある。具体的には、統計翻訳システムの精度向上のために、入力文に簡単な前処理（前変換）を行った上で、統計翻訳システムにかけることが試みられている。この前変換では、文の深い解析を行うのではなく、少数の規則で日本語の語順を英語の語順に近づけるといったことが行われる。この場合、変換後の「文」の人間から見ての可読容易性は保証されていない。

また、計算機処理に限定せずに産業と日本語との関わりを考えた場合には、分かりやすいマニュアルを作るための日本語の書き方の指導、といったことも、もちろん重要な事柄であろう。

3 産業と機械翻訳

日本語の計算機処理と産業の国際競争力という観点から、機械翻訳を取り巻く現状について眺めてみよう。欧米においては国際的企業が、従来から機械翻訳システムを活用している。この証左として、ここ数年の機械翻訳に関する国際会議（AMTA、EAMT、MT サミット）においては、ユーザトラックが設定され、機械翻訳の使用事例の報告がなされている。最近ではそのような国際的企業が日本の拠点において、日本語に対して機械翻訳の利用を始めている。この場合、外国でのサービスや製品の日本への導入ということになり、英日翻訳が行われることになる。おそらくは既に欧米言語間の翻訳に利用してきているということもあり、SYSTRAN や Language Weaver などの機械翻訳システムを使う例が多いようである。産業に限定しなければ、日常的には、日英であれ、英日であれ Microsoft や Google の機械翻訳を使う人も多いと思われるが、いずれにせよ、これらは全て海外企業の開発した機械翻訳システムである。

機械翻訳を用いた国際競争力の強化は現実的な話題であり、欧米では TAUS (Translation Automation User Society) という組織が機械翻訳を用いたビジネス翻訳の効率化を目指して積極的に活動している。製品のライフサイクルが短くなっている現在、乱暴な言い方をすれば、2年間のライフサイクルの製品のマニュアルを半年かけて翻訳すれば、4分の1のビジネスチャンスを失う。

機械翻訳が今、実用性を高めており、集中的な技術開発が必要であることはヨーロッパでは認知されており、たとえば EU の研究開発プロジェクトである FP7 において、以下のような機械翻訳関連プロジェクトが進行中である。

- EuroMatrixPlus
- "iTranslate4" (Internet Translators for all

- European Languages)
- META-NET (Multilingual Europe: a Technology Alliance)
- MOLTO - Multilingual Online Translation
- PANACEA (Platform for Automatic, Normalized Annotation and Cost- Effective Acquisition)
- PLuTO (Patent Language Translations Online)
- TTC - Terminology Extraction, Translation Tools and Comparable Corpora

上で述べたように、機械翻訳は既に実ビジネスで使われている。このようなビジネス文書の翻訳実務では、単純なテキスト翻訳の占める割合は少なく、既存文書の修正や活用が多いために、翻訳システム単体の能力ではなく、翻訳過程全体としてのサービスの提供が必要とされる。テキスト入力、対訳の提示（翻訳メモリ）、辞書引き、機械翻訳、フォーマット、翻訳メモリの管理、などを統合した環境を作ることが必要であろう。

また、最近の機械翻訳の研究開発の主流は用例翻訳や統計翻訳といった大量の言語データの存在を前提にしたものであり、利用できる言語データの収集が大きな問題となる。先に述べた TAUS では、TAUS Data Association (TDA) を設立し、データの共有を進めている。ここには、Adobe, CISCO, Dell, Intel, Microsoft, Oracle, Xerox といった企業も参加している。

4 我が国の現状

このような世界の趨勢と比べて、我が国の現状ははなはだ心もとないと言えよう。企業における自然言語処理の研究グループの縮小や、機械翻訳からの撤退などがあり、機械翻訳研究が衰退している。このため、企業において、用例翻訳や統計翻訳といった既存のルールベースから離れた手法のシステム開発し、製品化することができなくなっている。また、上に述べたような、研究開発

用のデータ共有においても、わが国の企業の参加は少ない。

これは単に自然言語処理技術応用産業の問題ではなく、我が国の企業の国際対応（多言語対応）への取り組みが不十分であることも原因であろう。自社の製品の開発サイクルの中に多言語対応を取り組み、機械翻訳などの情報技術を用いて効率化を図ることが、販路を広げ、ビジネスチャンスを広げることを強く認識すべきであろう。

本稿の初めに述べたように、情報処理技術の発展は、情報産業の競争力の強化に直結するとともに、情報処理技術をツールとして用いることにより、より広い産業における競争力の強化に資すると言われている。我が国の現状に当てはめてみれば、機械翻訳をはじめとする情報産業の競争力が低下し、ひいては、日本語情報処理能力の相対的低下による、我が国産業全体の国際競争力の低下をひきおこすことになる。

一般に、日本語の処理は英語などと比較して難しく、日本語からの、また日本語への機械翻訳の精度は他と比べて低いと言われてきた。統計翻訳においても、この傾向は続いている。もし、英語から他の言語へは、機械翻訳を用いて、高速・高精度で翻訳が可能であり、一方、日本語からの機械翻訳の精度が低いとすれば、英語から他言語への翻訳と比べて、日本語から他言語への翻訳は、コストと時間がかかることになり、我が国企業は製品情報の翻訳の速度、精度、費用で欧米企業に太刀打ちできないことになりかねない。

ところで、従来、日本語の特質と、それに基づく解析精度の低さに起因して、日英翻訳は英日翻訳より精度が低いと言われてきたが、大規模なデータを用いる統計翻訳においては、出力側の言語データの大きさの効果により、日英翻訳のほうが性能が良い場合があるようである。ここには「性能」とは何か、という問題も存在するが、本稿ではそこには立ち入らず、日本語に関わる翻訳においても他言語のデータの影響があるということに触れるに留める。

さて、ビジネス文書に対する翻訳システムの競争力の

向上が、産業全体の競争力の向上に繋がることから、わが国においても、テキスト翻訳システムの研究開発を進め、翻訳実務全体を高品質・高効率化する環境を開発することが重要である。また、いろいろな場面で多くの人に実際に機械翻訳を使ってもらうことが、多言語でのコミュニケーションにおける機械翻訳システムの有効性を実感してもらうためには必要であろう。学校や自治体の多言語サービスの中での機械翻訳システムの利用なども重要な活動である。

機械翻訳の有効性への認知が高まれば、企業での開発や利用が増え、当該分野のスタッフの数が増えることから、学生の就職先が増える。それが、この分野の学生数の増加になり、この分野の基礎研究の活発化に繋がる。こういった正のループが出来ることを強く望むものである。

5

産業日本語 プラットフォーム委員会

一般財団法人日本特許情報機構（Japio）では平成19年度から、産業日本語プラットフォーム開発に関するフィージビリティスタディを行ってきた。現在は特許版・産業日本語として、特許情報を中心に産業日本語に関する活動を進めている。具体的には、特許版産業日本語委員会を設置して、詳細な議論を進めている。また、AAMT/Japio 特許翻訳研究会において、特許の機械翻訳の精度向上に向けた研究調査を行っている。

これらの活動に加えて、今年度から、Japio のご支援を受けて、産業日本語研究会の実働部隊として産業日本語プラットフォーム委員会が活動を開始することとなった。自然言語処理の第一線の研究者と、企業からの参加者による構成で、ここまで述べてきたような観点から、産業と日本語とのかかわりを広く検討し、我が国の国際競争力を再び高めることを目指している。既に特許版産業日本語委員会や AAMT/Japio 特許翻訳研究会の活発な活動があることから、産業日本語プラットフォーム委員会は、これまでの Japio での議論を踏まえた上で、

より広い見地から（プラットフォームという語に制約されることもなく）、産業と日本語のかかわりについて議論することとしている。また、単なる議論に終わらないように、2年程度の活動期間を経て、具体的なプロジェクト提案にいたることを目標としている。

このプロジェクト提案では、情報処理技術によって、日本語の運用（生成、理解、検索など）を支援することを目指す。翻訳しやすい、検索しやすいという立場から見れば、利用しやすい日本語を作る環境を追求するともいえよう。

産業日本語プラットフォーム委員会は、この方向性の下、具体的な計画の立案と実行を進める。皆様のご指導・ご鞭撻をよろしくお願いいたします。

