



Pierce, J., Carroll, J., Hamp, E., Hays, D., Hockett, C., Oettinger, A., Perlis, A. : Language and Machines: Computers in Translation and Linguistics

also known as Report by Automatic Language Processing Advisory Committee (ALPAC),
National Academy of Sciences Publication 1416, Washington D.C. (1966)

コンピュータによって自動的に翻訳を行う「機械翻訳」の着想は、電子計算機の登場後まもなく、1946年にロックフェラー財団のW. Weaverによって唱えられた。その研究は1950年代に入って本格的にスタートし、1954年には早くも露英機械翻訳の簡単なシステムがジョージタウン大学とIBMの共同研究によって開発された。当時、アメリカ政府もスプートニク・ショックの影響から、ロシア語を中心とした科学技術文献の自動翻訳に強い興味を示し、巨額の研究費が投じられるようになっていった。第1の機械翻訳ブームである。

しかし、翻訳という非常に高度な作業がコンピュータによって簡単に実現できるわけではない。機械翻訳システムも、単純に単語を置き換えるものから、単語の活用、さらに文の構造を扱う、徐々に高度なものとなっていたが、それでもますます問題の難しさが認識されるということではなかった。

このような状況の中で、本当に実用的な機械翻訳システムが実現可能であるのか、機械翻訳研究の現状と将来を調査する委員会がアメリカ国立科学アカデミーに設置され、その報告書としてまとめられたのがALPAC報告書として知られる本書である（報告書自体は1965年に提出された）。結論は、近い将来に機械翻訳を実用化することは困難であり、機械翻訳の研究にかわり、言語の理解を目指す基礎的な研究を行うべきである、というものであった。この報告書を契機に世界的に機械翻訳に対してほとんど研究費が出なくなった。いわば、機械翻訳研究に暗黒の時代をもたらした報告書である。

報告書には、翻訳に対する社会的需要の調査、翻訳文の質の評価方法、翻訳システムを導入した場合のコストの試算、電子辞書を用いた人手による翻訳との比較などが、20項目の付録資料とともにまとめられている。機械翻訳についてさまざまな側面から徹底的な調査が行われたといつてよい。

さて、その後の機械翻訳の歴史をみると、ALPAC報告書の影響は大きかったものの地道な研究は継続され、1980年代には、多言語社会であるヨーロッパや、言語障壁に苦

しむ日本などで機械翻訳の国家プロジェクトが行われ、第2の機械翻訳ブームとなった。そこでは、言語の構造をもう少し正確に扱ったり、意味の問題を<人間>、<動物>、<植物>、<抽象物>などのレベルで近似することにより、翻訳の規則を詳細化するということが行われた。しかし、それはいわば翻訳（する文章）が対象としているこの世界の複雑さを詳細に記述していくことに等しく、質的満足を得る前にどこかで力尽きるということが繰り返された。

このような反省の中から、抽象化した規則を人間が与えるのではなく、翻訳された文章そのものを用いて翻訳を行うというまったく新しい発想が生まれた。現京都大学総長の長尾先生による1981年のTranslation by Analogyがその発端であり、さらに、IBMのグループが音声認識などで成果をあげていたNoisy Channel Modelの考え方からStatistical Machine Translationの定式化を行った。

このようなData-Drivenの機械翻訳は、その後しばらくは計算機パワーと対訳データの不足により十分に研究を発展させることができなかった。ところが、1990年代後半あたりから、ついに所望の計算機パワーが得られ、また情報化社会の中で電子化された対訳テキストが大量に利用可能となり、Data-Drivenの機械翻訳研究が活発化し始めた。インターネットによる多言語のビジネスチャンスや、安全保障上の問題から（紛争地域の情報を機械翻訳によって素早く入手したい）アメリカが再度、巨額の研究費を出すようになったこともこの状況を後押しした。

このようなわけで、現在、機械翻訳研究は3度目のブームを迎えている。機械翻訳は、計算機に知的なことを行わせるという「夢」への大きなテストベッドであり、また現実世界に存在する言葉の「壁」を解消する有望な手段である。そういう意味で何度ブームがきてもおかしくないのであるが、ブームに惑わされない冷静な視点を保つことも重要である。いまALPAC報告書を読み返してみるのはいよいことかもしれない（<http://www.nap.edu/books/ARC000005/html/index.html>で全文を電子的に読むことができる）。

(平成15年10月20日受付)

黒橋禎夫／東京大学大学院情報理工学系研究科
kuro@kc.t.u-tokyo.ac.jp

