



データ基盤から知識基盤へ

■ 黒橋 禎夫



20世紀の学術研究はさまざまな技術的革新をもたらしたが、同時に社会には環境問題、格差問題などの弊害も生まれ、人類はむしろ矛盾に満ちた世界で苦しんでいる。学術が、複合的な社会課題を解決し、真に人々の心の安寧をもたらすためには、人文・社会科学を含むさまざまな学術研究の協働が必要であり、それを可能にする土壌の構築が課題である。

これまでの20～30年を振り返ると、情報学が社会および学術界に対して広範なインパクトを与えてきた。およそ30年前に発明されたWWWは情報発信・流通の在り方を根本から覆し、社会構造を変革した。さらに、この10年あまりの深層学習の進展も目覚ましく、AlphaFoldによるタンパク質構造予測は生命科学研究に革命をもたらし、WWW上の膨大な対訳テキストから学習された機械翻訳システムDeepLは世界のコミュニケーションの形態を変容させつつある。最近では、ChatGPTが大学の定期試験等で合格レベルのエッセイを書くというニュースも駆け巡っている。このような情報学、そしていわゆるAIの進展が、引き続き今後の社会をドライブすることは間違いない。

21世紀の学術および社会の大きな潮流として、データの重要性が明確に認識された。さまざまな観測や計測からデータを作成し、デジタル化し、オープンにして議論・利活用することで学術的に大きな進展が起こっている。このような状況下で、我が国では国立情報学研究所(NII)を中心にSINET6に至るネットワーク整備と、研究データ基盤の整備が継続的に進められてきた。

さまざまな学術研究の協働を進める上で問題となるのは、特定分野の専門家も他分野については素人であ

■ 黒橋 禎夫
国立情報学研究所長／京都大学特定教授

1994年京都大学大学院工学研究科博士課程修了。博士（工学）。2006年より京都大学大学院情報学研究科教授。2023年4月より国立情報学研究所長。自然言語処理、知識情報処理の研究に従事。



り、多様な分野を見通してデータを直に活用することは容易ではないという点である。これからの学術研究が総合知として深化し、複合的な社会課題を解決していくためには、データの解釈、知識の関係付け・体系化を自動化し、分野を横断する新たな知の創造を支援する知識基盤の構築が必要である。そのような基盤の必要性は10年以上前から指摘されてきたが、データ基盤が整い、データをオープンにするものの価値が認識され始め、機械翻訳研究に端を発するAI基盤モデルにより論文やマルチメディアデータを高度に解釈することが可能となりつつあることから、ついに知識基盤の構築を本格的に目指すべき時代となった。しかし、AI基盤モデルの構築には大規模計算資源を必要とし、一部の海外企業による寡占化が進んでいることが大きな問題である。

このような急速な社会変革の時代に、2023年4月からNIIの所長を務めることとなった。私自身はこれまで自然言語処理の研究を行ってきた。深層学習とデータが注目されるはるか以前から、「文法ではなくテキスト（データ）が言語を説明する」という長尾真先生（指導教官）の影響を受け、言語解析ツールと言語コーパスの研究開発・公開に取り組み、計算機による言語理解というきわめて挑戦的な課題に研究コミュニティとして取り組むことの重要性を意識して研究を進めてきた。このような経験を活かし、今後は、我が国全体の大きな連携のもとに、AI基盤モデルの研究・開発・運用の体制を整備し、知識基盤の構築に全力で取り組んでいきたいと考えている。NII初代所長の猪瀬博先生の座右の銘が「ゆっくり急げ (Festina lente)」であることを知った。肝に命じたい。