

# 中国の研究事情

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) 前会長、  
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) 永くメンバー

✉ jtsujii@microsoft.com



## 1 はじめに

16年間勤務した東京大学を本年3月に早期退職し、5月からマイクロソフト研究所アジア (Microsoft Research Asia, MSRA) で研究生活をはじめている。日本の大学から米国の民間企業で、しかも中国にある研究所と大きく環境が変わり、適応に追われている。本稿では、5ヶ月という短い間であるが、その間に経験したMSRAや中国の研究事情について書く。

## 2 マイクロソフト研究所 アジアと中関村地区

MSRAは、正規の研究員だけで200人を超える研究所で、自然言語処理、機械翻訳、ウェブサーチといった分野の研究者だけでも40名、それに同数以上のインターン学生、滞在研究員がいるという大所帯である。また、米国シアトル郊外のレドモンドにある研究所、英国ケンブリッジ、米国シリコンバレーにある研究所や開発分門とも日常的に共同研究が進められており、非常に刺激に満ちた環境にある。

また、MSRAのある北京の中関村は、精華大学、北

京大学、中国科学院計算科学研究所、自動制御研究所、北京航空大、北京語言大学といった中国の代表的な大学や研究所が集積している文教地区である。精華大学や北京大学などの客員教授をしているMSRAの研究員もおり、MSRAの研究員がこれらの大学でのセミナーに参加したり、逆に、これらの大学からの研究者や学生がMSRAでの講演会に参加したり、物理的に近接していることもあってか、相互の交流は日本以上に活発である。

MSRAが中国の若手研究者を支援するプログラムを開始したことから、これらの大学からの若手研究者がMSRAに滞在研究員として数ヶ月滞在するのが恒常化している。

MSRAだけでなく、中国の研究機関は、海外の研究者との交流、とくに米国の研究者との交流が活発である。毎日のように外部からの研究者のトークが行われている。言語処理や情報検索の分野でも、カーネギーメロン大学のJ. カーボネル教授、楊教授、ジョンズホプキンス大学のK. チャーチ教授、メリーランド大学のD.W. オアード教授などのトークが、MSRAで開催された。

また、海外で活躍している中国や台湾出身の研究者との緊密な交流も中国全体に見られる傾向であろう。米国

の大学で活躍している中国系の研究者が、ひっきりなしに MSRA や中関村にある研究機関を訪問している。私が訪問した精華大学のバイオインフォマティクス系の学科でも、ほとんどが米国から帰国した研究者であり、米国の研究機関に所属しながら、同学科の客員教授や研究員をしている研究者も多数見られた。中国の研究機関は、在外の中国系研究者のネットワークをうまく活用して、急速に国際化している。

中関村には、先に述べた公的な研究機関だけでなく、北京大学からスピンアウトした企業（方正集团など）、中国での Twitter サービスをする Sena、少し離れるが、Google や百度といった多くの民間企業も集積している。若手の技術者や研究者の間ではこういった企業、研究機関、大学の間での移動も結構活発であり、昼食時には別の企業に移動した研究者の話などがよく話題にあがる。

シアトル郊外のレドモンド研究所やベルビューにあるマイクロソフトの検索サービス (Bing) 開発グループにも、別の大手会社から大量に人が移動してきている。このような技術者や研究者の移動は日常的に起こっている。

### 3 北京での国際学会

今年の7月には、北京で SIGIR という情報検索分野での国際会議が開催された。この分野での権威ある学会であり、850名程度の参加者があった。言語処理の国際学会である ACL、Coling といった会議も、急速に参加者が増加し、やはり、参加者が1000名近い会議になりつつある。ウェブ、電子出版、Twitter などのソーシャルメディアの普及など、テキスト情報の膨大な蓄積

への知的アクセスの必要性が認識されてきたことの反映であろう。

SIGIR での2つの基調講演は、いずれも表面上の単語や文字列に基づく検索から、それらがあらわす意味に基づく知的な検索へ向かう方向での講演であった。テキストマイニングと情報検索が急速に接近していること、また、情報検索から知識管理や知識処理に向かう強い傾向を示しており、興味深かった。個人的には、情報検索の高度化には、言語の意味や意味と知識の関係を扱う技術が不可欠であり、言語処理研究からの貢献が不可欠であると思っている。

具体的な研究内容とは別に、この SIGIR は国際的な舞台での中国の存在感を示すものとなった。北京という開催場所もあろうが、招待講演者2名も、米国で活躍しているとはいえ、中国系の研究者であり、発表者にも中国系の研究者が多かった。昨年度にやはり北京で行われた Coling、本年度米国ポートランドでの国際計算言語学会 (ACL) など、言語処理の主要学会でも同様な傾向であり、これら3つの学会では、発表論文の数で中国が米国に次いで第2位の座を占めるようになっている。米国など中国外からの論文でも第一著者が中国系留学生である論文まで入れると、発表論文の半数を中国人研究者が占めるようになっている。

また、SIGIR に先だって、国際計算言語学会 (ACL) と SIGIR との共催で Summer School が北京で開催されたが、200名の定員がすぐに満員になるなど、中国人学生の研究への意欲の高さを象徴するものとなった。国際学会も、中国での会員増加に熱心である。現在の ACL 会長の K. チャーチによると、会員の年齢構成をみると中国と日本の差が歴然としていて、中国の会員が非常に若いということであった。彼は、中国での学会員獲



得に熱心で、先の Summer School は、来年も定員を増やして開催する予定とのことであった。

ポートランドの ACL のあと、レドモンドのマイクロソフト研究所に10日間滞在した。言語処理や検索系の研究グループを訪ねたのだが、ここでも中国系の研究者が多く勤務していた。マイクロソフトのウェブ検索 (Bing) の開発責任者が中国系であることも影響しているのであろうが、開発部門だけでなく研究所でも中国系の研究者が目立つ。もちろん、レドモンドのマイクロソフトの研究所には、中国系だけでなく、インド系、ヨーロッパ系の研究者も非常に多く、米国生まれの米国育ちという研究者は圧倒的な少数派である。国境を越えて人材を集積できる米国企業の強さを感じた。

私の旧研究室でも、プロジェクトで雇用していた研究者は、米国、フィンランド、ノルウェイ、中国、韓国、ドイツなど多国籍化し、留学生を入れると、一時期10ヶ国以上の異なる国籍の人間が集まっていた。ただ、最近急速に変化してきているとはいえ、日本では海外からの研究者が、継続的な長い雇用の機会を得ることはまだ難しく、日本の大学の海外留学生が日本に定着しない傾向がある。この傾向は改善されてきてはいるが、日本社会の大きな損失になっていることは明らかだと痛感している。

## 4 意味検索

さて、私が現在興味を持っている研究について、書いておこう。ACL の会長講演で K. ナイトも「次世代の研究のキーワードは、Semantics、Semantics、Semantics だ」と強調していたように、やはり意味に

まつわる技術が、次の言語処理や検索技術の中核となる。

人工知能研究や言語処理研究が始まった初期のころから、意味 (Semantics) を処理する技術は重要な技術として認識され、実際、多くの研究者が取り組んできた。人工知能の初期の (主として MIT 系の研究) の研究を集め、当時の人工知能研究の方向に大きな影響を持った本のタイトルも、Semantic Processing であった。ただ、意味という、目に見えないもの、しかも、個人での差が大きく、不確定なものを扱うための技術が整わず、意味に関する研究は長い間、アイデアの段階にとどまり、実際の場面で使える耐性の高い技術にまで成熟することはなかった。

この困難がここ数年でおおきく解消されてきている、非常に大規模なテキストデータや、個々の環境下でのユーザの行動履歴、クエリとその後の URL クリック対の履歴などが計算機処理できる形で蓄積されるようになったことで、これまで抽象的で漠然としたものと考えられてきた意味が、大量のテキストデータやユーザの行動履歴から抽出できるようになったのである。また、意味と同様に漠然としていた知識も、ウェブ中のデータを処理することで、大規模な形で集積できるようになってきた。

このことは、多くの研究者をテキストの意味理解、その意味に基づく知的検索へと向かわせている。また、ユーザの行動履歴が大量に蓄積されるようになったことで、検索のための表面上のクエリに潜む、隠れたユーザの意図も取り出せるようになってきた。このクエリに潜む検索意図に合致するようなテキスト (あるいは、ウェブサイトを) を検索すること、この検索をテキストの意味に基づいて行うことが、私の当面の研究目標である。もちろん、このための基礎技術 (文やテキストからその意味

を取り出す技術)は、検索だけでなく、高品質な機械翻訳システムといった、これまで達成不可能と考えられていた技術をも可能にする。

言語、意味、知識という、トライアングルの謎解きをする事、この目標に向かって研究を進めていきたいと思っている。

## 5 おわりに

本稿では、中国の肯定的な側面を強調した。もちろん、中国社会には否定的な側面も多い。これについては、また稿を改めて書きたいと思っている。