言語情報処理への期待

最近，言語処理の研究は，大規模なコーパスを利用して，数理的処理をおこない，言語が使われる状況でのなまのテキストをもとに言語のもつ規則性や言語の統計的特徴を抽出するアプローチが大きい流れとなっている．機械翻訳でも，パラレルコーパスといわれる複数言語の対訳のテキストから，対訳辞書や対訳表現を検索評価する方法が，従来の規則をベースとした方法の問題点を解決できる部分が大きいとして，多くの研究機関で研究されている．これまでは，コーパスのもつスタテイックな言語的性質が主に利用されてきたが，今後は，これに加えて言語の使われる環境にダイナミックに適応するモデルを構築するなどの発展が期待される．

コーパスをベースとした言語処理の方法のメリットとして，言語的規則の作成を人手による作業から解放してコーパスから自動的に知識獲得できることが期待できること，共通のコーパスをもとに，解析，変換，生成をはじめ翻訳などの言語処理における方法のちがいを定量的に評価できること，人間がさまざまな知的活動を表現して多様な変化がみられる言語表現をそのままの形で扱えることなどがあげられる．

しかし，ここで問題となるのは，大規模なコーパスをどのように収集するかである．コーパスの規模はある程度以上の大きさがないと，言語的情報の抽出が安定におこなわれないといわれている．コーパスの作成には，多大の労力と費用を必要とする．また，どのようなコーパスを作成するかということは，どのような言語処理を対象としているかに関連が強く，コーパスの作成自体が研究であるともいわれる．そこで，日本語を中心とした言語処理の研究のために使用できるコーパスの作成と蓄積を研究者が共通に利用できる体制の整備が切望される．コーパスは人類共通の知的財産の一つと位置づけられるものであるので，言語処理の研究者の知恵を集めてコーパスの充実へ努力することが必要であろう．これにより，言語情報処理の研究が一層進展するものと確信している．

樽松　明
電気通信大学
電子工学科