



自然言語処理

野村 浩郷†

言語は永遠である。計算機も永遠である。しか
らば、言語の計算機処理は永遠であらねばなら
ぬ。もちろん、人の世が存続するかぎりにおいて
意味をもつ話しではあろうが！

自然言語処理の終局の目的は、人間がもち得る
言語活動の能力のすべてを計算機処理という形で
シミュレートし、それを人間社会において効果的
に活用することである。活用の形態は、言語表現
を媒体とするコミュニケーションである。活用の
目的は、人間社会における精神のおよび物質的な
豊かさである。効果的な活用は、コミュニケー
ションの経済性すなわち効率化である。言語の計
算機処理は、人間の社会の中でそれらを実現する
ためのものである。

言語は人間が備えるべき機能の本質の一つであ
り、言語を通じて、社会生活に欠くことのできな
い、思考、伝達、交渉、記録などを含むコミュニ
ケーションが可能となる。それゆえに、自然言語
処理は、拡大することはあっても終わることがな
い情報化社会において欠くべからざる要素であ
る。そして、最も重要なことであるが、常に新し
いことを見いだし創りだす人間社会の英知を集約
する壮大なるロマンでもある。

ことばの遊戯は、ロマンの顕著な一つの現象で
ある。このロマンは、遊びには留まらず、あると
きは文化を生みだし、あるときは社会に挑戦し、
そしてあるときは歴史を動かす。自然言語処理
は、人と計算機とが調和する社会へむかってのそ
のようなことばの遊戯を支援する建設的な歩みで
ある。技術の一步は小さいかも知れないが、技術
の積み重ねはロマンを現実のものとし豊かな社会
へと歩ませる大きな原動力である。

満ちたらんとしたい未来の社会において、ロマ
ンを科学たらしめ、そして技術たらしめるための

あくなき追究をさしおいて、外に何があるであろ
うか。ロマンは、余裕がある時代のお遊びでもな
ければ、悲壮な時代の妄想でもない。それへの取
り組みは、いかなる時代においても、きたるべき
堅実なる社会への基礎を築く果敢なる挑戦であ
る。言語は、好むと好まざるとにかかわらず、そ
の使命を担った人類の財産である。

自然言語処理は、言語の機械処理である。その
機械とは、いわゆる電子計算機である。電子計算
機が出現して以来、自然言語処理は、計算しない
計算機、すなわち思考する計算機の一つの姿とし
て活発に研究されてきた。そのような背景の中
で、今から約 30 年前に、言語学と計算機科学と
の融合による一つの大きなエポックがあった。自
然言語の文の内部構成のとらえ方の一つであるい
わゆる句構造に根ざした文法理論と、計算機のモ
デルであるオートマトンの理論および言語の形式
モデル化の一つである形式言語の理論とに裏付け
されて、文脈自由型と呼ばれる規則形式の文法記
述とそれに基づく句構造の構文解析技術が提案さ
れた。これは、英語などの言語族を対象とする現
在の自然言語処理技術にもそのままの姿で生きて
いる。その後、今から約 20 年前には、いわゆる
人工知能における初期の技術の一つとして、構造
表現形式とそれに基づく推論をとまなうパターン
マッチングによる意味構造解析技術が生まれた。
これは、いわゆる格文法と呼ばれる自然言語の文
法理論と結びついて、日本語を対象とする現在の
自然言語技術にそのままの姿で生きている。これ
とは独立に、同時期に、いわゆるカナ漢字変換技
術の基礎が確立され、現在の日本語ワードプロ
セッサへと発展していった。そして、今から約
10 年前には、これらを集約した形で、機械翻訳
システムという応用情報処理システムへの実用化
が胎動を始めた。また、これとは独立に言語の意
味を厳密に形式化するための新しい考え方の文法

理論が提出され始めた。これらの変遷を経て、現在は、能力はまだきわめて初歩的ではあるが実用化が達成され、機能とデータを整備して能力を段階的に向上させることと、使える機能を社会生活においてうまく活用することが精力的に追求されている。

コミュニケーションは、言語表現を媒体として概念を伝達する。概念は、言語表現が表す意味の人間による解釈である。伝達される概念は、客観的事実であったり、意図であったり、要求であったりする。したがって、自然言語処理は、概念のモデル、言語のモデル、および理解のモデルが要素となる。これらは、現在までの自然言語処理では、それぞれ、概念関係および概念構造、辞書および文法、解析および生成の方式として研究され、その具現化として計算機処理のための具体的なデータとプログラムが作成されてきた。

自然言語処理は、おおざっぱにいうと、言語学と心理学と計算機科学とからなる境界領域の科学・技術である。それぞれのモデルは、このような観点から考察されてきた。しかし、古くからある概念モデルとしてのシソーラスは、自然言語処理の技術を飛躍させるには至っていない。10年前ごろに提案された言語モデルとしての新しい文法理論は、10年が過ぎた今日、まだ実際の応用システムで力を発揮するまでには至っていない。20年前ごろにいろいろな提案があった理解モデルとしての意味理解機構は、20年が過ぎた今日、自然言語処理システムの設計の基礎にはなっていない。このような歴史から単純に推測すると、生物化学あるいは脳物理などの外の領域の研究がこれらの問題をまったく違う方法で一気に解決しないかぎりには、自然言語処理は永遠に着実な歩みを淡々と進めることになる。しかし、その過程では、技術は確実に進歩し、その結果として程度の差こそはあれ部分的な効果をもたらす。すなわち、自然言語処理はそのような高度な内容をもった対象であり、究極のロマンは一日にしては成らずとも身近なロマンはバベルの塔のブロックを一つずつ積み上げるように実現できる。

人間の社会において言語の発展が知能の発展とともにあったように、自然言語処理の発展も計算機の発展に支えられてきた。モチベーションはロマンにあったが、実現は計算機技術にあった。

辞書や文法のような言語知識と構文解析のような処理知識が、計算機の発展により量の拡大と分類の詳細化と対象の多様化として展開されてきた。量の拡大は、辞書の語数や文法の規則や処理の工数の増加である。分類の詳細化は、構文的・意味的な範疇や関係の細分化である。対象の多様化は、句から文へそして文から文脈へなどの拡張である。また、言語のマジックともいえる語用、すなわち熟語、比喩、照応、省略などを組み込む拡張である。さらに、言語表現の外にある心的状況や外部状況や知識の情報を取り込む拡張である。現在は、それらを試行錯誤する段階にある。10年後は、それらをモデル化する基礎を見いだす段階に達する。20年後は、それを形式化する基礎を見いだす段階に達する。30年後は、それを技術につなげる基礎を見いだす段階に達する。もしも、これらの時期が早まることがあったとしたら、それは研究者・技術者の優れた努力のたまものである。もしも、これらの時期が遅れることがあったとしたら、それは問題の深淵がより深刻であったことの発見である。その後は、実社会で役立つ技術の段階的な実現および改良と、新しく生じる言語現象への対処とが永遠に続く。その過程では、計算機処理技術に基づくさまざまな技巧が考案され、そして部分機能を実現する応用システムが社会に提供され、それを効果的に活用する方法が明らかにされる。

言語は、音声表現および文字表現の両方をもつ。言語表現には、時間を超えて伝達される静的なもの、実時間で交わされる動的なものがある。言語表現は、非言語表現によって代替される場合もある。視覚、聴覚、臭覚、触覚などがその例である。これらを言語表現化する問題もあれば、その逆の問題もある。また、これらに並行した言語表現の問題もある。

コミュニケーションは、人間の間だけにおいてみられる現象ではなく、人間とともに地球を共有する動物や植物、そして鉱物の間とでも行われるかも知れない。さらに、地球上における対象のみとだけではなく、宇宙のすべてにおける対象とも行われるかも知れない。人間がもつ言語は、言語の種類が異なっても、非常に類似な様相をもつ。動物・植物の言語もそのようなものかも知れない。宇宙も地球につながっている。言語こそ

が、地球をそして宇宙を平和に導く要素の一つである。

言語の習得のメカニズムは、言語理解の根元である。感情・感性・韻律・語用などは、言語とコミュニケーションを豊かにする。自然言語処理は、実質効果が高くかつ波及効果も大きい。自然言語の将来は、平坦ではないが明るく、夢に満ちている。言語のロマンは新たなロマンを生む。それゆえに、言語は尽きることがないロマンである。



野村 浩郷 (正会員)

1944年生。1967年大阪大学工学部通信工学科卒業。1969年同大学院修士課程修了。同年電電公社通研入所。1988年九州工業大学情報工学部教授。1972年工学博士。神経回路網、学習理論、知能処理、自然言語処理、機械翻訳などの研究に従事。



情報処理機械から情報記憶機械へ

牧之内 顕文†

1. 現在の計算機の使われ方

先日、銀行間の金の振込に関して、ある銀行に問い合わせる必要が生じた。そこで、振込用紙の写をつくづく眺めると、振込方法には「電信」と「文書」があり、問題の振込は「電信」で行ったとある。以前から私はこの種別に合点いかなかった。というのも現在の銀行業務はほとんど計算機化され、銀行の支店間はもちろん銀行間の金の移動は計算機間のデータ通信によって計算機内の帳簿（ファイル）上のデータの変更によっていると私は思い込んでいるからである。そうであれば振込はすべて「電信」によるものであるはずだ。

このように私達が日常接しているサービスはすでに計算機を抜きにしては語られなくなっている。銀行のサービス、JRの緑の窓口や航空機の座席予約サービス、さらにはNTTの電話番号案内（これが電算機化されてサービスの質がいかによくなったかは104番愛好者の私にはよく分かる）、宅配便の貨物追跡（かつて一度、宅配で送った貨物の所在を問い合わせたことがあった。すぐにその貨物の状況が分かったのには驚いた）などいろいろ思い当たることがあるだろう。これら社会シ

ステムは社会サービスの基盤に根をはっているためいったん故障すると大問題を引き起こす。銀行システムのダウンは必ず新聞記事になるのはこのためである。

さて、これら社会システムにおける計算機の使われ方を見ると、計算機のたとえば π の値をいかに精度よく早く計算できるかといった「計算能力」よりもむしろ「記憶能力」が使われていることが分かる。銀行勘定システムであれば預金者の口座への金の出し入れの記録、座席予約システムであれば座席の予約状況の記録が計算機内に記憶される。これら記録データは、量が膨大であることと記憶の安全性・永続性のために、二次記憶装置に蓄えられる。計算機の記憶能力とはこれら膨大なデータを効率よく覚え、忘れずかつ必要ときに必要なデータを思い出させる（検索する）ことができる能力のことである。

今、情報処理研究者で電子メールを利用していないのは皆無であろう。この電子メールは計算機の記憶能力とデータ交換能力を利用したものである。計算能力とはあまり関係ない。また、ワープロは会社の業務で使われているばかりでなく家庭でも使われ始めた。これもまた計算機の記憶能力に負うところが大きい。もし、作成された原稿が記憶されなくて修正や再利用ができない場合を考

データベース・システム研究会主査

†九州大学工学部情報工学科