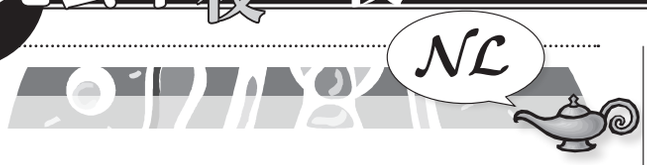

自然言語処理(NL)研究会

中川裕志　東京大学
森　辰則　横浜国立大学

自然言語処理研究会の沿革

自然言語処理研究会はNL研と通称しています．1975年に計算言語学研究会として設立され，1980年までこの名称でした．1981年に自然言語処理研究会と改名し，現在に至っています．20年を上回る歴史を持つ老舗の研究会です．歴代主査は，以下の方々です（敬称は略させていただきます）．

初代：和田　弘，第2代：長尾　真，第3代：吉田　将，第4代：田中穂積，第5代：野村浩郷，第6代：新田義彦，第7代：松本裕治，第8代：島津　明，第9代：中川裕志

現在，国会図書館長を務めておられる長尾先生をはじめとして，歴代主査の多くの方々はいまなお自然言語処理研究の第一線で活躍しておられます．華やかに脚光を浴びることは少ない地味な研究分野という印象を持たれる方も多いと思いますが，歴代主査の研究活動の息の長さからも分かるように，粘り強く研究を進展させている分野といえます．とはいえ，自然言語処理とは馴染みのない方も多いと思いますので，自然言語処理という学問分野自体の歴史を少し述べさせていただきます．

自然言語処理研究の歴史とソシュール

人間の使う言語についてはソクラテスの時代から考察され，研究も進められてきています．古くはアリストテレスの論理学も言語との関連が深いものです．この時代は，モノには正しい名前があるという思想が強かった時代でした．つまり，まずモノがあって，それに言語表現が付加されるという考え方だったわけです．その後，ローマ帝国の崩壊とともに科学は衰え，暗黒の中世を迎えます．ただし，暗黒の中世の間ですら，僧院の中で文法についての考察が進められ，17世紀のポールロワイヤル文法では形態素論，構文論なども現代のものに近い形で形式化されています．イギリスではロックが意味論に対応する概念さえ提唱していました．しかし，現代のように計算機で言語を処理することに哲学的基礎を与えたのは，言語学におけるコペルニクス的展開の提唱者と言われたソシュールです．その内容は，端的に言えば，すでに述べた「まずモノがあって，それに言語表現が付加される」という従来の考え方の否定から始まります．ソシュールは，言語について学問するには，言語をモノとの関係から切り離して，言語表現を独自の存在とみなし，それらの間に成り立つ関係を体系的に捉えることを提唱しています．さらに言語の歴史的発展を議論するのではなく，現時点における言語使用（これを言語の共時態と言います）を分析すべきと説きます．これは，計算機の内部での記号処理として自然言語処理を考える現在の考え方にほかなりません．すなわち，機械翻訳にせよ，文書分類，文書要約などという自然言語処理の成功した応用はすべて，言語の体系内部でのテキスト変換作業です．機械翻訳は異なる言語の間でのテキスト変換，分類，要約や言い換えは同じ言語の中でのテキスト変換であります．これらは，言語の外の世界，すなわち先ほど述べたモノとのかかわりは議論していません．まことにソシュールこそは，現代の自然言語処理の精神的祖といえるでしょう．

現代の自然言語処理技術

電子計算機が出現した1940年代後半から早くも自然言語処理の研究は開始されています．IBMのLuhnが1950年代に文書の自動要約の研究発表をしていますが，これはすでに現代の文書要約と類似の方法論を使っています．その後，言語学の成果を計算機に取り込むという方法論での研究が続きました．この流れは最近まで続き第6代主査の時代まで主流でした．しかし，第7代主査の時代から大量のテキストデータから統計的機械学習によってテキスト処理する方法が主流になります．その結果として，非常に精密な数理モデル化が重視される時代となって現在に至っています．振り返りますと，ソシュールの提案した言語の共時態のモデル化という考え方は，言語学者の発達させてきた枠組みの中から脱皮して，計算機で大量テキストの統計処理をするというパラダイムシフトを起こしたのがこの20年の変化と考えられます．

NL研の現在

NL研の現在の活動についてお話します．現在は年6回の研究会開催が主な活動です．発表件数も多く，1回

研究会当日に行われた学生奨励賞の表彰式の模様

の研究会が多くの場合 2 日にわたります．年間発表件数は 100 件を越える盛況な研究会です．研究内容的には，古典的な言語学を基礎とするもの，大量の言語データ，辞書データの扱いに関するもの，統計的機械学習に関するものが拮抗しています．

10 年前には多くの研究が行われたテキストを単語分割し品詞情報をつける形態素解析の研究は一段落しました．形態素解析の上のレベルの処理である構文解析はいつも NL 研の中心的テーマであり，現在もそのような流れは変わりません．応用としては，なんといっても機械翻訳が一番手ですが，文書分類，文書要約，自然言語インタフェースといった実用的技術の発表も続いています．少し将来的な話では質問応答がよく研究されているようです．なお，日本が世界に先駆けて提案し先導的立場を築いた技術としては，用例ベース機械翻訳と，言い換えの研究が特筆されるでしょう．

一方，対象とする言語は，日本語が一番多いのは変わりませんが，英語のほかに最近では中国語を扱う研究が大幅に増えました．言語処理の研究とは世界情勢に大きく影響されることが実感されます．

実は，これらの流れは，自然言語処理のトップレベル国際会議である ACL，COLING などでもほぼ同様です．Web で生き残る言語は最終的には英語と中国語になるといわれますが，自然言語処理の分野では，国際会議で対象とする言語はすでにほとんど英語と中国語だけになってしまいました．

数理的な仕掛けの方向を眺めますと，統計的機械学習でより数理的に精緻なモデルを使う方向への勢いはとどまるところを知りません．そういった意味では，長らく言語学の傘の中で行ってきた研究が，画像や音声と同じ数理的レベルに到達した感じがします．とはいえ，NL 研の参加者はやはり言語が好きな人が多いのです．数学と言語のハザマで日夜研究に励むというのが昨今の NL 研で発表をする研究者の姿ではないでしょうか．

若手へのアウトリーチ

さて，NL 研では，研究者の高齢化がそれとなく感じられるようになっていますが，分野の発展を考えれば，やはり若い人たちに自然言語処理に興味を持ってもらい発表してもらうことが大切と考えます．そこで，NL 研では，本年 5 月の研究会において学生セッションを設け，表彰もいたしました．その活動について次に述べます．

5 月の研究会は，音声言語情報処理研究会（SLP 研）と合同で開催され，学生セッションは両研究会による合同企画でした．卒業論文研究や修士論文研究等で得られた成果を研究会で発表していただき，ディスカッション等を通じて学生の皆さんの研究を振興するとともに，両研究会のさらなる活性化を促すことが本企画の趣旨です．特に，優れた研究発表には「学生奨励賞」を進呈し，表彰をすることを計画しました．一般の研究発表も含めた発表募集に対して，学生セッションとしては 7 件の発表がありました．内容としては，自然言語処理，音声言語情報処理の分野に関する幅広い話題があり，いずれも興味深い発表でした．表彰対象者の審査については会場にお越しの参加者の皆さんに広くお願いしました．審査用紙を事前に配布し，各々 7 段階の総合評価をしていただきました．集計作業は学生セッションの終了後直ちに行われ，総合評価の平均値に基づき，両研究会の幹事が協議をし，最終的に 2 名の発表者を受賞対象に決定しました．表彰式は，研究会の最後の時間帯に行われ，和やかな雰囲気の中で受賞者に表彰状等が贈呈されました．学生セッションならびに学生奨励賞の表彰は当研究会としては初めての試みでしたが，皆様のご協力をいただきまして，つつがなく執り行うことができました．今後も同様の企画により，研究会の振興に努めていきたいと考えています．末筆になりますが，学生セッションの開催につきまして，当日ご参加の皆様，ならびに，SLP 研の主査，幹事の皆様に多大なるご協力を賜りました．心より感謝を申し上げます．

（平成 19 年 7 月 4 日受付）

中川裕志（正会員）
nakagawa@dl.itc.u-tokyo.ac.jp
東京大学情報基盤センター教授．自然言語処理の研究に従事．最近は統計的機械学習に興味を持つ．2007 年度より本会 NL 研主査．用語抽出システム：言選 Web，用例表示システム：Kiwi などを公開．

森　辰則（正会員）
mori@forest.eis.ynu.ac.jp
横浜国立大学大学院環境情報研究院教授．自然言語処理の研究に従事．特に，質問応答や自動要約等の文書に対する情報アクセスに興味を持つ．2004 年より本会 NL 研幹事．