

## 第14回計算言語学国際会議 (COLING-92) 報告

林良彦\*1 石崎雅人\*1 古瀬蔵\*2 木下聡\*3

NTT 情報通信網研究所\*1

ATR 自動翻訳電話研究所\*2

東芝研究開発センター\*3

E-mail: hayashi@nttnlz.ntt.jp, ishizaki@nttnly.ntt.jp, furuse@atr-la.atr.co.jp, kino@isl.rdc.toshiba.co.jp

1992年7月23日から28日にかけて、第14回計算言語学国際会議 (The 14th International Conference on Computational Linguistics, COLING-92) がフランスのナント市会議場で開催された。COLING は2年に1回開催されている計算言語学、自然言語処理関係の最大の国際会議である。今回は、500人以上の参加者を集め、約420件の投稿から採択された約200件の論文発表が行なわれた。本報告では、会議の概要および、いくつかの興味ある論文を紹介する。

## A Report on COLING-92

Yoshihiko HAYASHI\*1, Masato ISHIZAKI\*1, Osamu FURUSE\*2, Satoshi KINOSHITA\*3

NTT Network Information Laboratories\*1

ATR Interpreting Telephony Research Laboratories\*2

Toshiba R&D Center\*3

E-mail: hayashi@nttnlz.ntt.jp, ishizaki@nttnly.ntt.jp, furuse@atr-la.atr.co.jp, kino@isl.rdc.toshiba.co.jp

The 14th International Conference on Computational Linguistics (COLING-92) was held at City Congress Hall of Nantes, France from July 23rd to 28th. COLING is the largest international conference in the field of computational linguistics and natural language processing. Over 500 people from all over the world attended the conference and about 200 papers were presented. This report gives the outline of the conference and summarizes several papers, which the authors are interested in.

## 1 はじめに

1992年7月23日から28日にかけて、第14回計算言語学国際会議(The 14th International Conference on Computational Linguistics, COLING-92)がフランスのナント市会議場で開催された。COLINGは、1965年の第1回以来、ほぼ2年に1回開催されている計算言語学、自然言語処理関係の最大規模の国際会議である。今回は、500人以上の参加者を集め、約420件の投稿から採択された約200件の論文発表が行なわれた。本報告では、会議の概要および、筆者らが興味を持ったいくつかの論文を紹介する。

## 2 COLING-92の概要

### 2.1 日程およびPreCOLING-92

本会議は、1992年7月23日から28日にかけて、日曜日を除く5日間開催された。また、これにさきかけて7月20日から22日まで、PreCOLING-92としてチュートリアルが開催された。筆者らは、チュートリアルには参加しなかったため、その詳細は不明であるが、「多言語間の移植性」をメインテーマとして以下の3つのトラックから構成された模様である。

- 多言語ソフトウェア(コードセットや多言語OS)
- 多言語ドキュメント(SGML<sup>1</sup>およびTEI<sup>2</sup>)
- 多言語辞書(多言語間での辞書交換を行なうための標準データ形式)

### 2.2 開催地

フランスのナント市会議場で開催された。ナントは、パリの南西約350km、大西洋岸に近いロワール川沿いに位置する人口約25万人的大都市である。歴史的には、1598年8月13日の「ナントの勅令」<sup>3</sup>で有名である。会場となった市会議場(Cite des Congres de Nantes)は、今年新設されたばかりの真新しい会議場であった。

### 2.3 参加状況

前回フィンランドのヘルシンキで開催されたCOLING-90[中村90]においては、約540人の参加があった。今回の会議で配布された最終参加者リストによれば、今回のCOLING-92も前回と同様の535人の参加があった。10人以上が参加した国別の参加者状況を図1に示す。地元フランスが1位であり、前回1位であったアメリカが2位であった。日本からの参加者は、前回から若干増えて今回は3位となった。最近の景気後退にも関わらず

<sup>1</sup>Standard Generalized Markup Language

<sup>2</sup>Text Encoding Initiative

<sup>3</sup>カトリックと新教徒との間に起こった宗教戦争を鎮圧するために、新教信仰の自由を条件付きながら認める勅令をアンリ4世が発した。

日本からの参加者が多いことは、依然として日本における自然言語処理、計算言語学に対する興味の高さを示していると考えられる。

## 2.4 論文投稿・査読の状況

前回のCOLING-90では、約600件の投稿論文から約200件が採録されたという。今回は、約420件の論文が投稿されて206件の論文が採録された。前回と同様に論文は、Topical paperとProject noteの2つの種別に分類された。両者の区分は明確ではないが、後者はシステムなどの実現に関する論文としてデモを原則として行なうことを要請された。

今回の論文募集では、マルチリンガルを推奨するというプログラム委員会の方針によって英語以外での投稿も歓迎された。その場合は、英語でのアブストラクト1枚を付加することが要請された。実際には、本文がフランス語で書かれた論文が数件見られた程度であった。

論文の採否は、以下のように決定された。まず、プログラム委員長のZampolli教授(イタリア・ピサ大学)がICCL(International Committee on Computational Linguistics)のメンバと協議して、計算言語学を9つの分野に分割し、それぞれにコーディネータを割り当てた。各分野のコーディネータの出す分野内のランクを重視して論文の採否が決定されたが、分野に渡って採否が問題となる場合は、デモを伴う論文が優先された。

9つの分野とそれぞれのコーディネータを表1に、採録された論文の分野ごとの分布を図2に示す。この分類では、Applicationsの件数が多い。ただしこの中には、機械翻訳、文書作成支援、情報検索といった応用の他にコーパス・事例に関する処理や文章生成の論文も何点か含まれているので注意が必要である。

なお、5件以上の論文が採録された国別の分布を図3に示す。ここでも日本が上位に入っている。

表1 9つの分野とコーディネータ

分野	コーディネータ
phonology, morphology, syntax	H.Uszkorcit
semantics, pragmatics	E. Hajicova
discourse, dialog	C. Sidner
computational methods	M. Liberman
tools	M. Kay
large-scale resources	D. Walker
applications	M. Nagao
NLP & hepermedia	S.G. Pulman
language industry	A. Ognowski

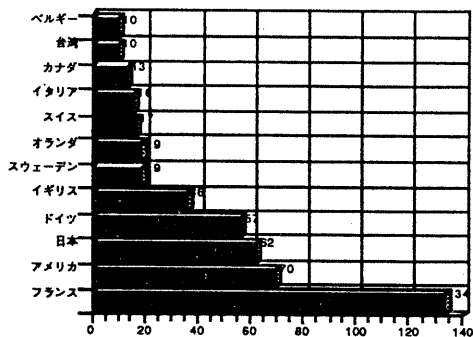


図1 参加者の国別分布

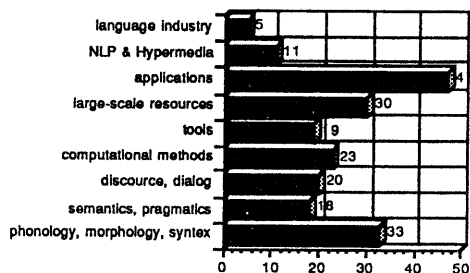


図2 発表論文の分野別分布

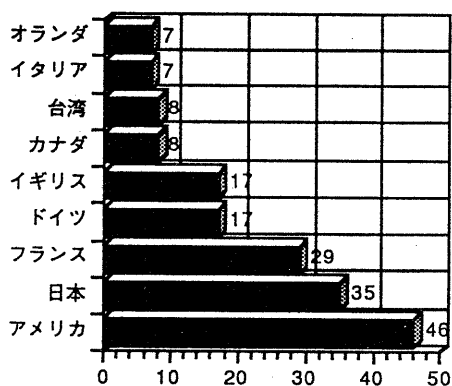


図3 発表論文の国別分布

## 2.5 会議の構成

会議の構成の骨子は、以下のとおりである。

- 招待講演：Saarbrücken大学のWahlster教授がマルチモーダルな通訳機 Verbmobilプロジェクトに関する招待講演を行なった。内容の概要は3節に示す。
- デモなし論文：3つのパラレルセッションによって発表が行なわれた。
- デモつき論文：多くのProject noteといくつかのTopical Paperについては、デモが行なわれた。デモ付き論文の発表時間は、毎日昼休みの時間帯に設けられた。1件の発表ごとに1つの小部屋が用意され、発表者は主に用意されたSun-WS、IBM-PC、Macなどを用いてデモを行ないながら発表を行なった。端末を囲んでの熱心な議論が行なわれた部屋もあったが、発表者がいない部屋もあったのは残念であった。発表する側は大変な面があるものの、イメージをつかみやすく、かつ、親密な雰囲気での議論ができることから良い企画であると思われる。
- パネル討論：クロージングを含めて4つのパネル討論が企画された。内容の概要については、5節で紹介する。
- 国際プロジェクト発表：計17件の各国で行なわれている国家的プロジェクトの紹介が行なわれた。詳細は、会場で配布された[Varile92]にまとめられている。

なお、セッション構成の特徴としては、聴講者がパラレルセッションを効率良く回ることができるよう、並行するセッション中の各論文発表が同期して行なわれるよう論文間にバッファの時間が設けられるとともに、発表者には時間を厳守することが要求された。パラレルセッションの運営方法として優れたものと思われる。

## 3 招待講演

“Computational Models of Face-to-Face Dialogues: Multimodality, Negotiation and Translation”, Wolfgang Wahlster

1991年にドイツが国家プロジェクトとして始めた Verbmobilに関する講演である。このプロジェクトで目指すシステムは、母国語が異なる人間同士のFace-to-Faceでの対話を支援する、一種の通訳システムである。従来提案されてきた翻訳電話のようなシステムとの大きな違いは、その利用形態である。すなわち、このシステムでは、機械に翻訳させる部分を利用者が指定するようになっているのである。例えば、日本人とドイツ人が英語を使って会話をする場合には、日英翻訳と独英翻訳を行う2台

の翻訳機(Verbmobil)をそれぞれが使用する。利用者は、通常、自分自身で直接英語を話して対話するが、英語で表現できない場合(例えば単語が思い出せない場合など)に、翻訳機を使用することになる。また、翻訳結果が、聞き手だけでなく、話し手にも聞こえるようになっていくという設定は、話し手が翻訳結果を聞いて、翻訳の妥当性をチェックすることを前提としてシステムを使用することを示唆するものである。

このような利用形態は、自然言語処理が目指す目標としては、翻訳電話などに比べ、明らかにトーンダウンしていると印象を受けざるをえない。すなわち、Face-to-Faceでの対話という状況設定は、言語による情報伝達を、ジェスチャーや顔の表情などと並ぶチャンネルの1つであると捉えていることを意味しており、言語表現だけに依存する翻訳電話などに比べれば、翻訳精度は相対的に低くてもすむとも取れるのである。もちろん、好意的に見れば、この先10年後程度(このプロジェクトが終了するのは2001年である)の近い将来における翻訳システムの精度や処理速度を考えた場合には、実現可能性が高い目標設定は評価できるし、マルチモーダルな対話という枠組の中で、自然言語処理を捉え直すことで、新たな知見が得られることも期待できよう。

## 4 主な論文発表

以下では、筆者らが興味を持ったいくつかの論文を紹介する<sup>4</sup>。なお、コーパス・事例を用いた処理に関しては、既に網羅的な報告がなされている[野美山92]ので、そちらを参照して頂きたい。括弧内は紹介担当者を示す。

### 4.1 基盤技術関連

#### 4.1.1 解析、生成関連

- “Probabilistic Tree-adjoining Grammar As a Framework for Statistical Natural Language Processing”, Philip Resnik, pp.418-424., および

“Stochastic Lexicalized Tree-Adjoining Grammars”, Yves Schabes, pp.425-432.

いずれの論文も、70年代に提案された構文木導出操作の文法Tree Adjoining Grammar(TAG)の統計的手法による拡張について述べている。ResnikはProbabilistic TAG(PTAG)、SchabesはStochastic Lexicalized TAG(SLTAG)とそれぞれの枠組を呼んでいるが、基本的な考え方は同じである。PTAGもSLTAGも、部分木は必ず語彙を含み、substitutionあるいはadjoiningという操作で木の結合を行う。そして、木の結合確率を付与することによって語彙依存性と階層的構文構造を実現する。Resnikは統計的TAGの有効性や問題点を

<sup>4</sup>これらの論文の選出は筆者らの興味に偏っており、会議全体の中から重要なものを抜粋したのでは必ずしもないことを御了解願いたい。

明解に説明しているが、確率パラメータの付与アルゴリズムや具体的な検証については言及していない。それに対し、Schabesは、文の生成確率の計算アルゴリズム、各規則の確率パラメータを推定するInside-Outsideアルゴリズムなどについて、SLTAGを詳細に形式化している。また、人工言語および、Air Travel Information Service(ATIS)コーパスの話し言葉の品詞列について確率パラメータの推定実験を行ない、SLTAGが言語を正しくモデル化し、文法規則を構築する際、Inside-Outsideアルゴリズムによって確率パラメータをStochastic CFGに比べ速く収束できることを示している。(古瀬)

- “A Fast Algorithm for the Generation of Referring Expressions”, E.Reiter and R.Dale, pp.232-238.  
心理言語学の知見、対話のトランスクリプトの分析結果を採り入れて、定名詞句の修飾語とheadの名詞をどのように選定すべきかを論じ、聞き手の認識を容易にする立場から対話における指示表現を高速に生成するアルゴリズムについて述べている。対話者が指示において好む属性のリスト順に、聞き手が認知でき、指示物を限定するのに貢献するなどの条件により属性値を選定していき、その属性値情報をもとに指示表現を生成する。このアルゴリズムは、on-lineドキュメント生成を行なうIDASシステムにインプリメントされており、指示表現のほとんどを修飾語の数が0か1の名詞句で表現している。(古瀬)
- “Acquisition of Selectional Patterns”, Ralph Grishman and John Sterling, pp.658-664.  
従来行われているコーパスからの学習で多いのは、学習エラーをなくするという観点から、間違いないデータを学習するというものだが、ここで提案されている学習では、曖昧性のある解析結果の情報も利用されている。例えば、文を解析してN個の解析木が得られる場合に、各解析木から抽出したデータに対し、単純に1/Nの重みをつけて学習する方法や、統計的文法(もちろん、学習用のテキストとは別のトレーニング用テキストを使ってチューンしたものを)を用いて解析し、その上位M個を学習するといった学習方法を取っている。これらの方法を使って行った学習の結果は、人間が抽出したデータと比較し、方式の有効性を確認している。ただし、今回の学習データは、動詞と、その表層格およびフィラーの3項関係であり、このデータを組み合わせて、複数の格を持つ格フレームにする部分は今後の課題となっている。(木下)
- Neumann, G. and G. van Noord, “Self-Monitoring with Reversible Grammars”, pp.700-706.

Neumann と van Noord は、解析生成両用文法を利用して、生成した結果を解析して多義がある場合に再生成を試みる機構を提案している。単純な生成検査法の非効率性を改善するため、トップダウンに多義性を生みだすノードを探索し、必要のある部分だけを再生成する。また、決定されたノードでうまくいかない場合には1つ上のノードを試すようにしている。解析と生成を関連付ける研究は少なく、貴重な試みではあると言うことができる。ただし、多義性をこのように解析生成の協調によって解決するか、内容決定と表層文生成のインタフェースで解決するかについては議論の余地のあるところであると思う。(石崎)

#### 4.1.2 対話研究関連

- Carletta, J., "Planning to Fail, Not Failing to Plan: Risk-taking and Recovery in Task-oriented Dialogue", pp.896-900.

Carletta は、地図の経路に関する情報提供対話を分析して、不完全な情報下で発話する場合 (risk-taking と呼ばれる) と不完全な情報を明確にしながら発話する場合があることを指摘している。この区別をプログラムとして実現するために、タスク、談話構造、表層文生成、プランニング過程についてのパラメータを設定している。また、対話データの分析により、risk-taking な主体のプランニング実行失敗時の回復方策として、パートナーの持つゴールの採用、パートナーへのターン委譲、話題の詳細化、情報省略、繰り返し、問題の無視、プラン前提条件の再充足、再プランニングがあると主張している。プランニングの過程を推論の対象とする必要性から、計算機構として Stefik が提案したメタプランニングの枠組を応用している。彼女のプログラムは、プログラム同士が対話するようになっており、ロボット間のコミュニケーションなどを応用として考えるのもおもしろいかもしれない。(石崎)

- Walker, M.A., "Redundancy in Collaborative Dialogue", pp.345-351.

Walker は、Pollack と Hirshberg が収集したラジオトークショーにおける対話を分析し、情報的に冗長である「繰り返し」表現に、相互信念の仮説を指示する証拠を提供する機能、言及されている命題を焦点化する機能があると主張している。今回の論文では前者について検討している。彼女は、単純なプランニングモデルにおける4つの前提-(1) 命題は信じられているかいないかの2つの状態でしか評価されない、(2) 主体のアクションの効果は達成されたとして文脈に加えられる、(3) 主体は論理的全知であるとされる、(4) 談話文脈は関係付けられていない-を、対話データを

用いて批判している。これを基にして、相互信念が証拠に基づいて推論される、棄却可能な仮説であること、及び、話し手は自分の発話行為の効果聞き手の応答によりモニターしているということとを主張している。

Carletta, Walker の研究は、対話コーパスの対話の分析、対話データからの単純なプランニングに対する批判から成っている。第1の点に関しては、彼女らの観察は必ずしもオリジナルなものではないが(それぞれ、Shadbolt, Clark and Schaefer という先行研究がある)、分析を計算機構のなかに位置付けたこと、及び分析が領域知識ではなく、対話に関する知識であるという意味で一般性を持つことが新しいと思う。(石崎)

- Eugenio, B.D. and M. White, "On the Interpretation of Natural Language Instructions", pp.1147-1151.

Eugenio は、命令を解釈して実際に実行するためには必要に応じて仮説を作りながらプランニングを行なう必要性を主張している。アイディアは目新しいものではないが、この研究がその一部を成している、自然言語によりアニメーションを駆動することを目標とする AnimNL プロジェクト (Pennsylvania 大学) は、マルチメディア時代の自然言語処理の一応用として興味深い。(石崎)

- "Balkanski, C.T., Action Relations in Rationale Clauses and Means Clauses", pp.267-273.

Balkanski は、命令文によく見られる手段節や理由節の解釈手順を提案している。彼女の研究は、AnimNL プロジェクトの一部ではないが、自然言語とプラン言語との関係の明らかにしようとしているという意味でその基礎を成すものとして捉えてよいと思う。彼女は、プラン言語の正しさの基準として自然言語で表現される行動間の意味関係、代数関係を利用している。このような正当性をサポートしようという努力は、応用面からみるとありがたみが少ないかもしれないが、対話研究のように方法論が確立していない分野では大切なことではないかと思う。

今回の Eugenio, Balkanski の研究は、Winograd の SHRDLU を思い起こさせる。見方によれば、彼女達の研究をより複雑な行動命令文の解釈についての検討であると考えることができる。これらの研究が SHRDLU からどのくらい進歩したと考えるかは評価がわかれるところではないかと思う。(石崎)

## 4.2 応用関連

### 4.2.1 機械翻訳関連

- “A New Quantitative Quality Measure for Machine Translation Systems”, Keh-Yih Su 他, pp.433-439.

この論文では、MT が出力した翻訳結果を、定量的に評価する新しい評価方法を提案している。従来の評価方法では、訳文の「正確さ」や「わかりやすさ」というような、評価者の主観に大きく依存する評価基準を用いていた。そういった評価基準も重要であることは明らかであるが、評価者が原文や訳文を読む必要があるため、評価に時間がかかる上、評価者によって評価のゆれが生じるといった問題がある。ここで提案された方法は、MT の翻訳結果をユーザが満足する訳文に直すのに、編集コスト（例えば、語句の挿入や削除、移動等）がどれくらいかかるかという点から評価するというものである。したがって、評価で使用する目標訳を前もって用意しておく必要があるが、上記のような評価時間の問題や評価者の違いからくる評価のゆれの問題はないわけである。

もちろん、この方法が、従来の評価方法と単純におき替わるとは考えられない。しかし、文法や辞書の修正が翻訳結果にどのように影響したかを短時間で知りたい MT 開発者にとって有用だけでなく、MT の導入を考えているユーザが、複数の MT システムの性能を比較する場合にも有用であると主張している点には注目すべきであろう。（木下）

- “Learning Translation Templates from Bilingual Text”, Hiroyuki Kaji, pp.672-678.

タイトルにある通り、対訳テキストから翻訳用テンプレートを学習するという話である。興味深いのは、テンプレートの作成にあたって、原文と訳文の構文構造の対応づけを、システムが自動的に行っている点である。すなわちこの手法では、対応する語および句構造の対応関係を、原文と訳文を、チャート法を使って同時に解析しながら取っつき、その過程で、原文もしくは訳文が持つ曖昧性も解消されるのである。

最近、日本を中心として EBMT（実例に基づく機械翻訳システム）の研究が盛んになっているが、これまで提案されてきたの多くは、こういった対応関係づけがなされた対訳データベースが存在することを前提としているにも関わらず、それをどうやって構築するかに関しての研究はあまり進んでいなかったように思われる。ここで用いられている対応づけの手法は、このようなデータベース

の作成にも有効であると思われ、応用が期待される。（木下）

- “Hierarchical Lexical Structure and Interpretive Mapping in Machine Translation”, Teruko Mitamura, Eric H Nyberg, pp.1254-1258.

入力文の詳細な意味理解によって高品質の翻訳を行なう Knowledge-Based Machine Translation(KBMT)を実現するための、階層的な語彙知識構造について述べている。筆者らは、語彙の構文属性と意味属性についての関連性について着目し、語彙知識のレベルを5階層に構造化している。そして、階層での inheritance を利用して、大規模な語彙知識の獲得、語彙知識の冗長でない記述、効率的な機械翻訳処理を目指している。本論文で述べられている階層的な構造は、CMU の英日の KBMT システムで動詞の語彙知識の記述に利用されている。なお、本会議のデモンストレーションセッションにおいて KBMT システムのビデオによる紹介が行なわれた。（古瀬）

### 4.2.2 その他の応用

- “A Case Study of Natural Language Customisation: The Practical Effects of World Knowledge”, Marilyn A. Walker, Andrew L. Nelson, Phil Senton, pp.820-826.

既存の自然言語インタフェース (NLI) を DB 検索の特定の応用にカスタマイズする方法を提案している。使用した NLI は、質問入力から DB 検索言語生成を行うモジュール、意味表現からの生成モジュール、演繹推論を行うモジュールを持つ。また、ドメイン知識として概念階層を持っている。対象ドメインは、セールスに関する DB である。

カスタマイズの手法は、以下の4ステップからなる。まず、ユーザのシステムへの質問のトランスクリプトを分析する。次に、DB 設計者によって与えられる E-R 図 (Entity-Relation Diagram) をドメインの概念表現と対し、トランスクリプトに現われるユーザの用語を E-R 図中のエンティティまたはリレーションへマッピングする。第3のステップでは、注釈付けられた E-R 図とテストスイートを入力としてカスタマイゼーションファイルを構築する。カスタマイゼーションファイルには、ドメインに依存する応用についての知識とドメインに独立な言語/世界知識に関係付けが記述される。これは、英単語と DB リレーション、属性、属性値とのマッピングを指定すること、および、推論モジュールによって用いられる概念を指定することにより行われる。最後のステップでは、テストスイートを新たに生成してカスタマイズの評価を行う。テストスイートは、トランスクリプトの分

折から定められた要求を満たす文を公表されているテストストから選択した文集をもとに生成される。以上の過程を繰り返すことにより、インクリメンタルなカスタマイズが可能であるとしている。

筆者らは評価を通して、(1)E-R図がドメインにおける語の使用を十分に表現しきれない場合があること、(2)NLIが具備している意味モデルをカスタマイズ者が直接修正できることが望ましいことなどを主張している。後者の主張は、外付け的な扱いによるカスタマイズには限界があることを意味するという点で否定的な結果であるが、NLIの実現においては考慮すべき主張であろう。(林)

- “Design Tool Combining Keyword Analyzer and Case-based Parser for Developing Natural Language Database Interfaces”, Hideo Shimazu, Seigo Arita, Yousuke Takashima, pp.735-741.

DB検索の自然言語インタフェースを実現する際には、処理効率および非文法的な入力に対するロバスト性が重要であるとの考えから、事例に基づくパーサ(CBP)とキーワード解析(KBP)を組み合わせた自然言語インタフェースの実現方式を提案している。

システム的设计者は以下のような過程により、自然言語インタフェースを構築する。まず入力のコバスをCBPにより解析する。CBPはあらかじめ蓄積された事例ベースを検索し、入力の言語パターンに対応するDB質問言語(SQL式)を生成する。この方法では、あらかじめ登録されていない言語パターンには対処できない。このような場合には、KBPが起動される。KBPは、あらかじめ登録されたキーワードとDBモデルとの関係からSQL式を生成する。このSQL式が質問に対して適切でない場合、言語パターンと正しいSQL式を事例ベースに登録することになる。一方、構築されたシステムでは、まずKBPを適用しこれが失敗した場合のみCBPを適用する。

以上のように本アプローチの特徴は、非文法的な入力に対応できるKBPとより強力な能力を持つCBPを相互補完的に用いる点にあるが、評価データが一切示されていないため、構築されたインタフェースの能力は不明である。特に、どの位の頻度で事例ベースへの新規登録が必要になるか興味あるところである。(林)

- “From COGRAM to ALCOGRAM: Toward a Controlled English Grammar Checker”, Geet Andriens, Dirk Schreurs, pp.595-601.

知的なテキスト処理や機械翻訳をより現実的なものとするために、制限言語を用いるという考え方

がある。筆者らは、まずAECMA,IBM,Ericssonの3社において技術情報のコミュニケーションのために設計された制限英語を分析した。その結果から、これらの制限英語は、言語学的基盤に乏しく、非文法性と制限範囲外の区別も明らかでないといった問題点があるとしている。

以上の分析に基づき、筆者らはCOGRAMと呼ぶ制限英語を開発した。これは、最大8000語の語リストと150規則を持つ文法からなる。文法は、語彙、統語、スタイルの3つのコンポーネントからなる。プロトタイプ版の机上の評価によれば、制限範囲のチェックだけでなく、一般的なミスを発見するのにも有効であったが、完全な制限を行なうにはまだ不十分であったという。

論文ではさらに、COGRAMを計算機で処理するための文法ALCOGRAMの概要について説明している。また、テクニカルライティングのCAIへの応用と、文法/スタイルチェッカへの応用の構想について述べている。特に後者については、InterleafのDTP環境へ組み込み中であるという。制限言語の話題は、古くて新しい話題であるが、これを支援する実用的なツールはあまり存在しない。本ツールの実フィールドでの評価が待たれる。(林)

- “Towards Computer-aided Linguistic Engineering”, Remi Zajac, pp.827-834.

本論文では、再利用可能な言語仕様の自動構築と自然言語処理プログラムの自動生成に基づく「コンピュータ支援言語工学」と呼ぶ枠組を提唱している。基本的な主張は、処理すべきテキストに現れる言語現象こそ言語処理プログラムの仕様を与えるものであり、現在利用可能な技術を結果することにより目的に応じた自然言語処理プログラムが(半)自動生成できるというものである。

言語の仕様記述には、タイプ付き素性構造TFSを用いる。言語学者がコバサ解析ツールを用いて仕様を記述していく。仕様記述においては、オブジェクト指向の考えに基づくことにより抽象化・詳細化が可能である。また、インクリメンタルに仕様を開発していくことができる。一方、ターゲットのプログラミング言語としては、同じくTFSをデータ構造とする制約論理型プログラミング言語LIFEを採用しているため、仕様は実行可能である。目的に応じた処理プログラムを生成する際には、適切な制御情報を加えることになる。これにより言語仕様の汎用性が保たれるとともに、目的に応じた効率の良いプログラムの生成が可能となる。本論文で述べられた枠組は、すでにESPRITやEUREKAといったプロジェクトで用いられたつあるという。

本論文では基本的な枠組を示すという段階にとどまっており、詳細な内容は示されていないが、野心的なプロジェクトであり、今後の成果が期待される。(林)

## 5 パネル討論

以下の4つのパネル討論が企画された。

- 研究費に関する国際協力 (Moderator: A. Zampolli) : 自然言語処理、音声処理の R & D 活動を促進する研究費についての国際協力について議論された。これらの技術の健全な発展のためには、限られた処理能力を有効に活かす良い応用を探ることが重要であることが強調された。この点から、機械翻訳以外では、テキスト処理、情報検索への適用が有望であろうとの考えが示された。
- 言語産業 (Moderator: H. Karlgren)
- 再利用可能な言語データ (Moderator: M. Liberman) : この2つのパネルでは、コーパスデータや文法などを協力して蓄積し共有の財産とすること、また、共有のコーパスを共通の評価の土俵とすることにより、協調的競争を行なうことの重要性が協調された。
- クロージング (Moderator: A. Zampolli) : プログラム委員会で制定された9つの分野の各コーディネータから各分野についてのまとめが行われた。

## 6 おわりに

今回の会議の印象を羅列すると、以下のようになる。

- 新しいパラダイム・研究の流れをリードする発表はなかった。
- しかし、自然言語処理をより実用的なものとするための、ロバスト性・効率向上のための研究が盛んである。
- 特に前者の問題については、コーパスや用例に基づく言語知識獲得や統計的な解析手法の研究が盛んであり、ピークを迎えつつある感がある。
- さらに、コーパスなどの言語データの蓄積・利用に関して国際的な協力の重要性が強調されてきた。
- 機械翻訳以外の「良い応用」を探す動きが、文章作成支援や情報検索などを対象として見られるものの、依然として機械翻訳が応用の主流を占めている。

前回との比較をしてみると、前回は対話型機械翻訳が話題を集めたが、今回はほとんど話題にならなかった。今回は、コーパス・用例に基づく自然言語処理が盛んで

あったが、次回までにどのような発展を見せるか興味深いところである。また、次回へ向けては、斬新な方法論・パラダイムの提案とともに、応用面ではメディア処理との統合などを含む先進的な「良い応用」システムの発表が期待される。

なお、次回の COLING-94 は、京大の長尾眞教授をプログラム委員長として、京都で開催される予定とのことである。

## 参考文献

- [中村 90] 中村順一. COLING90 に出席して, 情報処理学会自然言語処理研究会 80-3, 1990.
- [Varile92] G.B.Varile, A.Zampolli (Eds). Synopses of American, European and Japanese Projects Presented at the International Projects Day at COLING 1992, *Linguistica Computazionale*, Volume VIII, 1992.
- [野美山 92] 野美山, 武田, 渡辺, 丸山. COLING'92 および FGSLP'92 の報告 - 事例を利用した自然言語処理を中心として -, 電子情報通信学会言語理解とコミュニケーション研究会, 1992.