

国際パーズング技術ワークショップ参加報告

松本 一則

小暮 潔

国際電信電話上福岡研究所

ATR自動翻訳電話研究所

元年8月28日より4日間、ペンシルバニア州ピッツバーグにおいて開催された国際パーズング技術ワークショップについて、その概要を報告する。このワークショップでは、10のテクニカルセッションにおける47件の発表の他に、カーネギー・メロン大学機械翻訳センタ(CMU/CMT)において、センタの紹介とデモが行われた。

The Report on the International Workshop on Parsing Technologies

Kazunori MATSUMOTO
KDD Kamifukuoka R&D labs
2-1-15, Ohara, Kamifukuoka-shi,
Saitama, 356 Japan

Kiyoshi KOGURE
ATR Interpreting Telephony Research Laboratories
Sanpeidani, Inuidani, Seika-cho, Soraku-gun
Kyoto, 619-02 Japan

The International Workshop on Parsing Technologies was held at Pittsburgh, August 28-31 1989. During the workshop, 47 talks were presented, as well as demonstration of Center for the Machine Translation at Carnegie Mellon University.

1. はじめに

自然言語処理の特にパーシングについて広く議論を行うために、第一回目のワークショップとして、昨年8月28日から8月31日の間、ペンシルバニア州ピッツバーグにおいて、国際パーシング技術ワークショップ(International Workshop on Parsing Technologies)が開催された。

ワークショップは、最初の2日間がカーネギーメロン大学近くにあるHoliday Innで、後半の2日間が郊外のHidden Valleyというリゾートホテルで行われた。発表件数は47件であった。

また、2日目の最後には、カーネギーメロン大学の機械翻訳センタ(Center for Machine Translation)において、ビームサーチを利用した不特定話者の音声認識システムの実演、オンライン英日機械翻訳システムのビデオによる説明、汎用巨大知識ベース(VLKBs)用知識獲得・保守ツールの紹介等が行われた。

2. 講演の概要

以下、発表の中から、報告者らの興味を持ったものを中心に報告する。

Unification and Classification: An Experiment in Information-Based Parsing

Robert T. Kasper, USC/ISI - USA

著者は、86、87、88年と続けてACLで単一化に基づいた言語情報記述の定式化と、その拡張(選言、否定、含意の利用)および実装を報告している。しかし、拡張された単一化は、実行速度の点で問題がある。今回、クラシフィケーション(新しいクラスやオブジェクトを包含関係順のラティスに入れる操作)に基づいたシステムによって、計算量が減らせる可能性があるとして指摘している。今後、著者はクラシフィケーションが行えるシステムLOOMで文法や知識表現を記述し、効果を確かめる予定である。

An Overview of Disjunctive Constraint Satisfaction

John T. Maxwell and Ronald M. Kaplan, Xerox PARC - USA

選言を用いた制約の充足可能性判定アルゴリズムと充足可能な時のモデルを生成するアルゴリズムを

示している。充足可能性判定は、 $(1) \phi_1 \vee \phi_2 \Leftrightarrow \exists p: (p \rightarrow \phi_1) \wedge (\neg p \rightarrow \phi_2)$ を用いて連言にする。 $(2) P_1 \neq P_2$ の時、 $(P_1 \rightarrow t_1 = t_2) \wedge (P_2 \rightarrow \phi)$ を $(P_1 \rightarrow t_1 = t_2) \wedge (P_2 \wedge \neg P_1 \rightarrow \phi) \wedge (P_2 \wedge P_1 \rightarrow \phi[t_2/t_1])$ への書換えを適用できなくなるまで行う。 (3) 書換えられた式の内、充足不能な式の仮定(連言)の否定の充足可能性の判定を行う。著者は、自然言語の文法は、文中の異なった部分から取り出した素性構造は、異なった部分が多いので、元の式より、変形後の式の方が充足判定が簡単であると主張している。

Head-Driven Parsing

Martin Kay, Xerox PARC - USA

1980年にチャート解析法を提唱した著者が、Head-Driven 文法からチャートパーザを生成する Prolog プログラムを説明している。また、アクティブチャート・パーザを構成するための概略を述べている

Parsing with Principles: Predicting a Phrasal Node Before Its Head Appears

Edward Gibson, Laboratory for Computational Linguistics, CMU - USA

最近の言語理論に基づいた解析を行う時の解析アルゴリズムでは、ヘッドが出現する前に、その句の投射を予測する仮説駆動アルゴリズムが必要になる。ここでは、Constructed Parallel Parser (CPP) で用いられた手法を紹介している。まず、語彙からその最大投射(C-節点)をつくる(例: [DetP [Det' [Det the]]])。次に各C-節点のすぐ右にヘッドが現れた時の節点(H-節点)を予測する(例: [NP [DetP [Det' [Det the]]] [N' [N e]]])。そして、次の語が入力され、specifier attachment や argument attachment が行われる際、この予測に適合した節点を得られるようになっている。

Probabilistic Parsing Method for Sentence Disambiguation

T. Fujisaki, F. Jelinek, J. Cocke, E. Black, IBM-T. J. Watson Center - USA, T. Nishino, Tokyo Science University - Japan

著者は、84年のCOLLING等で、意味的制約や談話上の制約が、コーパスから得られる統計情報で補強できることを主張している。ここでは各書換え規則がその適用確率を持つようなパーザの構成方法と、このパーザを英文と日本語適用した結果から元の文法が補強できたことを報告している。

Probabilistic LR Parsing for Speech Recognition

J. H. Wright and E.N. Wringley, University of Bristol, UNITED KINGDOM

各書換え規則に適用確率が与えてあるCFG(PCFG)に対する拡張LRパーザの構成手法をSLRとLALRの場合について説明してある。生成するLRテーブルでは、シフト、還元、シフト-還元の小動作ごとに確率が与えられる。この手法をマルコフモデルと比較すると、マルコフの方が、柔軟で利用しやすいが、過去の履歴情報が限られる欠点がある。著者は、単語認識率が悪く、入力文が長い時、この文法モデルの方がすぐれていることをグラフで示している。

Parsing Continuous Speech by HMM-LR Method

Kenji Kita, Takeshi Kawabata, Hiroaki Saito, ATR Interpreting Telephony Research Laboratories - JAPAN

拡張LR構文解析法における動作表を用いて音声データ中の音韻を予測し、予測された音韻の存在をHMM音韻照合モジュールで調べることにより、音声認識と言語解析を同時進行させる方式(HMM-LR法)を提案している。この方式では、音韻認識と言語処理の間に、音韻ラティス等の中間的なデータを介さないため、高精度かつ効率的な認識処理系を構成することができる。

Parsing Japanese Spoken Sentences Based on HPSG

Kiyoshi Kogure, ATR Interpreting Telephony Research Laboratories - JAPAN

日本語の話し言葉を語彙駆動型の文法を用いて解析する手法を提案している。日本語の話し言葉の重要な特徴である複雑な文末表現やゼロ代名詞の取扱いに重点を置いた単一化文法と、この文法に従

い、音声認識結果のラティスを解析する解析機構を用いる。この手法では、統語・意味の情報だけでなく、運用的情報を用い、曖昧性の解消を行う。

Probabilistic Method in Dependency Grammar Parsing

Job M. van Zuijlen, BSO/Research - THE NETHERLANDS

コーパスに基づいた解析をおこなう翻訳システムについて紹介が行われている。コーパス情報のデータベースには、文ごとに、好ましい依存構造の木表現(節点はガバナ)や、その部分木で対訳の対象となるもの(TU)が蓄えられている。解析は、まず入力単語を節点に持つTUの集合を求める。これらのTUを繋ぎ合わせてできる解釈可能な依存構造の木の集合の内、節点が入力単語と良くマッチするものを解析結果としている。また、ローバスタな解析等ではシミュレーテッド・アニーリング法はあまり役立たないことが説明されている。

Handling of Ill-Designed Grammars in Tomita's Parsing Algorithm

Rohman Nozohoor-Farshi, University of Windsor - CANADA

本来の富田アルゴリズムでは、 ϵ -規則を含むCFG規則が扱えなかった。オリジナルのグラフ構造スタックに巡回部分グラフを導入したことで元のアルゴリズムの効率を落とすことなく ϵ -規則の利用を可能にしている。

Connectionist Models of Language

James L. McClelland, Carnegie Mellon University - USA

多くの聴衆が集まり、招待講演的な感じで発表が行われた。発表者はコネクショニストの立場からは、言語処理は、制約を満たすプロセスとして見なせると指摘し、コネクションモデルが言語学者が考えるほど能力低い物ではないことを主張した。コネクションモデルでも、前回の出力を記憶し、これを中間層への入力とすることで状態機械の学習ができる等の話題を提供した。

A Connectionist Parser Aimed at Spoken Language

Ajay Jain and Alex Waibel, Carnegie Mellon University - USA

今まではコネクションネットワークによる文法の学習は困難であったが、独自のネットワークを用意することで学習を行っている。時間順に、単語が1語づつ入力できるように、時間の経過に伴い出力値が減少していくユニットや、学習効率を上げるためのゲートユニットを利用している。学習は、句レベルと格関係の2段階にわけて、行われた。学習対象は、語彙が40語の受身形を許す単文である、学習後、(1) 文入力の途中での格標識の予想、(2) 入力誤り、(3) 文入力の速度の変動、に良い結果が得られたと報告している。

Massively Parallel Parsing in Φ DmDialog: Integrated Architecture for Parsing Speech Inputs

Hiroaki Kitano, Teruko Mitamura and Masaru Tomita, CMT/CMU - USA

談話翻訳システム Φ DmDialogの解析機能を紹介している。ここでは、音韻レベルから談話レベルまで、DMA型のマーカバッシングに基づいて統一的に音声入力の解析が行われている。また、形態素レベルから談話レベルの知識の利用が行われ、コスト(仮説のコストはその元となった仮説のコストから計算されている)に基づいた曖昧性解消が実現されている。特に談話理解が言語モデルの能力の向上と音韻処理における予測として、役立っていると主張している。

Finite State Machines from Feature Grammars

Alan W. Black, University of Edinburgh - UNITED KINGDOM

属性文法を等価な有限オートマトンに変換する手法が示されている。著者は、実時間の音声認識への応用を考えている。実際にこの手法を実装し、連言を含んだ31個のGPSG風規則を約30分間で状態数約9,000のオートマトンに変換している。

An Effective Enumeration Algorithm of Parser for Ambiguous CFL

Tadashi Seko, Nariyoshi Yamai, Nara National College of Technology - Japan, Noboru Kubo and Toru Kawata, SHARP Corp., - JAPAN

以前は、ある限られた正則言語でしか、解析木を数え上げ式に効率良く生成できなかった。これに対し、著者は、任意のCFGに適用可能な効率よい数え上げアルゴリズムを示している。まず階層グラフ(パースグラフ)が提案され、このグラフを用いれば、受理したい順の数え上げが、このグラフ上で長さの順にパスを見つけることと等価になることが示される。そして、このグラフを用いた効率よい数え上げアルゴリズムが述べられている。n語の文からk個の解析木を、各解析規則に付けられた重みの総和の順に生成するのに、 $O(n^3+kn^2)$ の時間と、 $O(n^2+kn)$ の空間で、できることも示されている。

Chart Parsing for Loosely Coupled Parallel Systems

Henry S. Thompson, University of Edinburgh - UNITED KINGDOM

疎結合のマルチプロセッサシステムにおける並列構文解析のため、チャート法を自然に拡張することで、頂点を各プロセッサに割当て、弧を対応するプロセッサ間で通信する方法を提案している。この方法をHyper Cubeやネットワーク結合されたリスブマシン間のremote evalで実験している。並列化と通信量のトレードオフによって、弧当たりの計算量が大きくないと並列化の効果がでなかった。また、符号化しやすい不活性弧のみを他の全プロセッサに送信する手法も試みられているが、送信する不活性弧の増加により、効果はなかった。しかし、高いプロセッサ間通信の実現により、将来的には有望と述べてある。

Parallel LR Parsing Based on Logic Programming

Hozumi Tanaka and Hiroaki Numazaki, Tokyo Institute of Technology - JAPAN

一般化LR構文解析アルゴリズムに基づいた並列解析システム(PLR)について述べている。PLRは並列論理型言語GHC上に実装されており、LRテーブルの各エントリを1つのプロセスとみなし、シフト、還元の動作を行う。もしシフト-還元のコ

フリクトが起こったら、シフト、還元を実行するサブプロセスを走らせる。ここでは、グラフ構造スタックを用いず同期をとらない版が紹介されており、著者の最近の merge_stack の導入による木構造スタックの実現はなされていない。

Learning Cooccurrences by Using a Parser

Kazunori Matsumoto, Hiroshi Sakaki and Shingo Kuroiwa, KDD Kamifukuoka R&D Labs. - JAPAN

文例とその解析木の組から単語間の共起関係を蓄積、利用する手法を提案するとともに、その手法によって構文解析の性能が向上したことを実験で示したものである。

Parsing, Word Associations and Typical Predicate-Argument Relations

Kenneth Church, William Gale, Patrick Hanks, and Donald Hinle, AT&T Bell Labs. - USA

89年のACL等で同一文中に現れる単語の組の統計情報の利用性の高さを主張する著者が、AP通信のニュースをコーパスとして行った、ここ数年の実験結果をまとめて発表している。ある単語 x や y がおのおのコーパスに出現する確率と、 x と y が同時に出現する確率を基にして求めた相互情報量を尺度にして、各種言語現象を提示している。また、統語解析木と組み合わせて、drink の目的語を抽出している。こうした技術が、辞書編纂者や計算言語学者の有用なツールとなり、OCRや音声認識システムに役立つと主張している。

PREMO: Parsing by Conspicuous Lexical Consumption

Brian M. Slator and Yorick Wilks, New Mexico State University - USA

知識ベースに基づいたブリファランスを利用して意味解析を行うシステム PREMO が詳細に説明されている。PREMO では、解析途中の各部分解析結果が language object と呼ばれる構造で扱われる。PREMO の制御はOSの様な働きをしていて、language object のプライオリティ・キューで実現されている。このキューの先頭の language object が

もっともスコアが高いが、このスコアは文法予測、意味チェック、談話の一貫性に重みをかけるブリファランスによって計算している。他の language object は alive か、not actively parsed といった状態を与えることで、ベスト-ファーストサーチによる解析を実現している。OECDから語彙情報作られ、文法情報や下位分類情報の他に文法予測が含まれている。

Parsing Algorithms 2-Dimensional Language

Masaru Tomita, CMT/CMU - USA

2次元テキストの理解のために、CFGを自然に拡張した2次元CFG (2D-CFG) を紹介している。応用例として、OCRによる名刺やビジネス・レターの理解等をあげている。そして、2D-CFG を効率よく解析するためために、アリーのアルゴリズムと LR(0) 解析アルゴリズムの拡張版である 2D-Earley と 2D-LR のアルゴリズムを提示している。

3. 終わりに

以下、紹介しきれなかった文献を列挙しておく。

Using Restriction to Optimize Unification Parsing
Dale Gerdemann, University of Illinois - USA

A Uniform Formal Framework for Parsing
Bernard Lang, INRIA - France

Head-Driven Bidirectional Parsing: A Tabular Method
Giorgio Satta and Oliviero Stock University of Padova - Italy

The Computational Implementation of Principle Based Parsers
Sandiway Fong and Robert Berwick, MIT, AI Laboratory - USA

A sequential Truncation Parsing Algorithm Based on the Score Function
Keh-Yih Su, National Tsing Hua Univ., Jong-Nae Wang and Mei-Hui Su BTC R&D Ctr., Jing-Shin Chang, National Chiao Tung Univ. - CHINA

Parsing Speech for Structure and Prominence

- Dieter Huber, University of Gothburg -
SWEDEN
- Predictive Normal Forms for Composition in
Categorical Grammars
Kent Wittenburg and Robert Wall, MCC - USA
- Parsing Spoken Language Using Combinatory
Grammars
Mark Steedman, University of Pennsylvania -
USA
- Recognition of Combinatory Categorical Grammars
and Linear Indexed Grammars
K. Vijay-Shanker, University of Delaware and
David J. Weir, Northwestern University - USA
- Analysis of Tomita's Algorithm for General
Context-Free Parsing
James R. Kipps, The RANDO Corp. - USA
- The Computational Complexity of Tomita's
Parsing Algorithm
Mark Johnson, Brown University - USA
- Probabilistic Parsing for Spoken Language
Applications
Stephanie Seneff, MIT - USA
- Parallel Parsing Strategies in NLP
Anton Nijholt, Twente University of
Technology - THE NETHERLANDS
- Complexity and Decidability in Left-Associative
Grammar
Roland Hausser, Computational Linguistics
Laboratory, CMU - USA
- The Selection of a Parsing Strategy for an On-Line
Machine Translation system in a sublanguage
domain. A practical comparison
Patrick Shann, University of Geneva and
- A Morphological Parser for Linguistic Exploration
David Weber, Summer Institute of Linguistics,
CMU - USA
- The Parallel Expert Parser: A Meaning-Oriented,
Lexically-Guided, Parallel-Interactive Model of
Natural Language Understanding
- Geert Adriaens, Siemens NLP Research -
BELGIUM
- The Relevance of Lexicalization to Parsing
Yves Schabes and Aravind K. Joshi, University
of Pennsylvania - USA
- A Framework for the Development of Natural
Language Grammars
Massimo Marino, University of Pisa - ITALY
- An Efficient Method for Parsing Erroneous Input
Stuart Malone and Sue Felshin, MIT - USA
- Analysis Techniques for Korean Sentences Based
on Lexical Functional Grammar
Deok Ho Toon and Yung Taek Kim, Seoul
National University - KOREA
University of Zurich - SWITZERLAND
- An Efficient, Primarily Bottom-Up Parser for
Unification Grammars
Neil K. Simpkins and Peter J. Hancox, Aston
University - UNITED KINGDOM
- A Broad-Coverage Natural Language Analysis
System
Karen Jensen, IBM - USA
- Pseudo Parsing Swift-Answer Algorithm
S. Pal Asija, Shelton, Connecticut - USA
- A Dependency-Based Parser for Topic and Focus
Eva Hajicova, Charles University
CZECHOSLOVAKIA
- Parsing Generalized Phrased Structure Grammar
with Dynamic Expansion
Navin Budhiraja, Subrata Mitra, Harish
Karnick, Rajeev Sangal, Indian Institute of
Technology Kanpur - INDIA