

PACLING-II 報告

中村 順一 甲斐 郷子
九州工業大学 情報工学部
〒 820 飯塚市川津 680-4

あらまし

PACLING-II (The Second Conference of the Pacific Association for Computational Linguistics) が 1995 年 4 月 19 日から 22 日に、オーストラリア Brisbane の The University of Queensland で開催された。参加者は約 50 人で、日本、オーストラリア、韓国が中心であったが、その他、米国、カナダや英国からの参加者もあった。4 つのキーノート・アドレスと 35 の論文発表が行われた。筆者らは、これに参加したので、その概要について報告する。

キーワード

Report on PACLING-II

Jun-ichi NAKAMURA and Kyoko KAI
Department of Artificial Intelligence
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820, JAPAN

Abstract

PACLING-II (The Second Conference of the Pacific Association for Computational Linguistics) was held at The University of Queensland, Brisbane, Australia, from April 19 to 22, 1995. Total number of participants was about 50: many from Japan, Australia, and Korea, and some from U.S.A, Canada and United Kingdom. 4 keynote addresses and 35 papers were presented. Since the authors of this report attended this conference, we reviewed its outline.

key words

1 はじめに

PACLING-II (The Second Conference of the Pacific Association for Computational Linguistics) が 1995 年 4 月 19 日から 22 日に、オーストラリア Brisbane の The University of Queensland で開催された。参加者は約 50 人で、日本、オーストラリア、韓国が中心であったが、その他、米国、カナダや英国からの参加者もあった。4 つの keynote address と 2 つの平行・セッションによる 35 の論文発表が行われた。

Brisbane は、オーストラリア第 3 の都市で、Queensland の州都である。オーストラリア北部に位置し、気候は、亜熱帯性であるが、乾燥しており、非常に住みごごちのよい、美しい町で、治安も良い所であった。近くには、Lone Pine Koara Sanctuary というコアラやカンガルーを始めとするオーストラリア独特の動物を真近に見ることのできる公園など、自然にあふれた場所である。また、The University of Queensland は、114 ヘクタールという広大なキャンパスを持ち、やはり自然に囲まれた美しい所であった。

会議自体は、19 日の 9:15 から開始され、まず、オープニングで、九工大の岡田直之先生が、PACLING の歴史や、趣旨を説明された。なお、岡田直之先生は、PACLING の会長でもある。PACLING は、1989 年 11 月に、オーストラリアの Melbourne で開催され、1991 年 10 月に第 2 回が日本の飯塚市 (九工大) で開催された Japan-Australia joint symposia on natural language processing がその前身となっている。第 2 回の開催の際に、オーストラリアと日本だけでなく、環太平洋の諸国に対象を広げることが決定され、PACLING と名称を変更した。その第 1 回は、1993 年 4 月にカナダの Vancouver で開催され、今回は、第 2 回にあたる。

PACLING は、環太平洋地域での計算言語学に関する情報交換がその目的である。しかし、大規模なお祭りの会議ではなく、小人数によるアットホームな雰囲気大切にすること、若手研究者の積極的な参加を推奨し、次の世代をになう研究者を育てることが重視されている。今回もその趣旨は十分、生かされていた。

続いて、開催地である The University of Queensland の Research and Graduate Studies の Pro-Vice-Chancellor である Peter Sheehan 先生が、自然言語処理研究のサーベイを元にして、歓迎の挨拶をされた。以下で

は、4 つの keynote address の内容と、筆者らが興味を持った発表について紹介する。

2 Keynote Address

Charles FILLMORE (University of California at Berkeley)

Dictionaries in the twenty-first century

計算機可読辞書 (Machine Readable Dictionary) を中心として、今後の辞書のあるべき姿について議論があった。まず、辞書記述として、

A is left to B vs A is right of B

Pat is to the left of Mrs. Jones

という例で、単に left/right を「方向」というだけの従来の NLP 的な意味だけではなく、実際に動作するために必要な Robotic semantics という観点からも考える必要を示した。

次に辞書の役割を 1. Decoding, 2. Weak Encoding, 3. Strong Encoding, 4. Resource for automatic lexicon building という 4 つの観点から説明した。たとえば、String Encoding として、

A is to B as C is to D

Snow is to white as blood is to red

のように、他の語との比較で示してはどうか、との提案があった。Resource for automatic lexicon building については、一般辞書の定義文のパターン (especially, usually, often など) から、典型的なもの、一般的なものなど、概念間の関係をうまく抽出すべきである、との指摘があった。

M.A.K. HALLIDAY (University of Sydney)

Computing meanings: Some reflections on past experience and present prospects

最初に、NLP の歴史は、1950-65: Machine Translation before ALPAC, 1965-80: Artificial Intelligence, 1980-95: Natural Language Processing に分け考えるべきであることを示した。1960 年代は、“If you can't compute your grammar, your grammar is wrong.” と楽観的に考えていたが、1980 年代には、文法が複雑化し

たことに従って “How much do I need to simplify my grammar.” という疑問が提出された。そして、1995 年からは、Fuzzy Logic を用いた意味計算が必要であるとの提案をした。

Michio SUGENO (Tokyo Institute of Technology)

Intelligent fuzzy computing

Fuzzy Logic の紹介のあと、単語の意味や概念のあいまいさや、Feeling/Mind の漠然性を Fuzzy Logic で扱う手法の例の紹介があった。たとえば、単語の意味の Extension は明確でなく、Intension は、明瞭ではない。そこで、Membership Function と Fuzzy Proposition を用いることにより、ある程度、表現できる。

また、人間の脳のソフトウェアは、Fuzzy Logic 的に自然言語を処理しており、その結果、外部への出力は Symbolic Logic で扱える、というモデルの説明があった。更に、Wittgenstein の Language-game (the meaning of a word is its use in the language) モデルと Systemic Functional Grammar の類似性についても解説があった。

その他、Fuzzy Logic 関連の研究、たとえば、EEE (Everybody, Everywhere, Everything) Computing プロジェクトの紹介があった。

John BATEMAN (IPSI, Gesellschaft fuer Mathematik und Datenverarbeitung, Germany)
Why get a translation when you can say it yourself? – multilingual text generation and information systems

文章を多言語で生成する手法について解説があった。まず、文章生成の応用例として、情報洪水に対処するための Information Filtering が示された。自然言語生成技術を用いることにより、特定の聴者に適したように情報を加工したり、対話状況や対話のゴールにあわせて情報を提示できる、という長所の説明があった。

次に、言語生成の問題点として、1. difficult, 2. resource / labor / theory-intensive, 3. expensive を示し、多言語文章生成システムにとって、1. Re-usability for different tasks/applications, 2. maximal re-use of resources across languages, 3. realistic generation with large scale resources が重要であることが述べられた。

その後、現在 IPSI で進行中の ‘KOMET’ Project について、その目標 (to determine the features of multilingual resources appropriate for practical generation) や、生成の Framework、生成例の紹介があった。特に、Systemic Grammar の system のうち、chooser (network) は、multilingual に使える点が強調され、中国語や日本語の生成例が示された。

3 一般発表

以下では、筆者らが出席したセッションを中心に、発表論文を紹介する。

Ken BARKER and Stan SZPAKOWICZ (University of Ottawa)

Interactive semantic analysis of clause-level relationships

Clause 間の意味関係を、表層的な構文情報を用いて解釈するシステムの発表があった。Noun-Modifier, Case-frame analysis, Clause-level relationships (CLR) analysis の 3 つのステップ (モジュール) で解析する。if, unless などの接続詞に関して、モダリティ (e.g. may) など各種の情報を用いて、意味関係を分析する。曖昧さがある場合は、preference rule により判定する。The Ontario Building Code から取った 100 文についてテストし、この手法の有効性を調べた。

Yoshio ERIGUCHI and Tsuyoshi KITANI (NTT Data Communications Systems Corporation)

Information extraction from Japanese text using Tomita’s generalized LR parser

情報検索 / 抽出のニーズが高まっているが、従来は、パターンマッチングを用いることが多かった。この研究では、Tomita’s GLR を拡張した skip parser (要素をスキップできるようにする) を作成し、これを efficient pattern matcher として使用して、情報抽出を行う。対象は日本語のニュースである。形態素解析、固有名詞認識、パターンマッチング (Unification 付きの context free rule) の 3 ステップで、文節単位で処理を行う。文節単位で処理した結果、recall rate 56%, precision rate 76% の実験結果が得られた。

Francis BOND, Kentaro OGURA and Satoru IKEHARA (NTT Communication Science Laboratories)

Possessive pronouns as determiners in Japanese-to-English machine translation

所有代名詞の日英の使用方法の差 (e.g. 私は舌を噛んだ / I bite MY tongue, 自分の机で勉強しなさい / Study at YOUR OWN desk) を調査し, 既存の機械翻訳システムの翻訳精度を向上するための, ヒューリスティック・ルールを追加した, ヒューリスティック・ルールとしては, “名詞句の head が, 身体の一部, 仕事などであれば, 日本語では, 所有表現を用いない” といったものである. 実験結果として, accuracy が 81%, precision が 88% となった.

Geunbae LEE, Hanmin JUNG and Jong-Hyeok LEE (Pohang University)

Bi-directional memory-based dialogue translation: the KEMDT approach

Massive Parallel/Marker Passing による韓英双方向翻訳の試作についての発表があった. 基本的には, 北野 1990 のアーキテクチャを韓英に利用している. ただし, 韓国語における単語の省略や語順のバリエーションに対応するため, marker passing の手法を拡張している. Topicalization などは扱っていない.

Hyun S. PARK (University of Cambridge)

Handling of scrambling in Korean using MC-TAGS

Joshi の Tree Adjoining Grammars (TAGs) を拡張した Multi-Component TAG (MC-TAG) を用いて, 韓国語のスクランプリングを扱う手法について発表があった. 韓国語のスクランプリングには, local なも (clause 内) のだけでなく, long-distance なものがある. 前者は, 各動詞に対する引数の置換で扱えるが, 後者は, Dominance Link と呼ぶメカニズムで扱う.

Fumiyo FUKUMOTO (Yamanashi University) and Jun-ichi TSUJII (UMIST)

Word-sense disambiguation using the extracted polysemous information from corpora

単語 (動詞) に複数の意味がある場合を含んだ統計的な処理方法についての発表があった. たとえば, take は目的語として TIME / STATE / PRODUCT の 3 種類の異なったものを取るが, これらの意味の分離を行う. 手法としては Overlapping Clustering Method を持ちこた. Wall Street Journal から取った 192,992 文に関する実験結果の報告があった.

Alpha K. LUK (Microsoft, Sydney)

Sense disambiguation of frequently used polysemous nouns

頻出語である course, head, order, point, power, service の 6 語について, Lesk の word overlap approach を改良した Lexical Matching Sense Disambiguation Algorithm により, 意味の分離を行う実験結果について発表があった. これは, 使用された単語の前後の単語の辞書 (LDOCE) 定義文で共通に用いられている単語の頻度を計算するとにより行う. この場合, 辞書を多段に参照し (multiple level dictionary look-up), 精度を上げている.

Arne JÖNSSON (Monash University, Sweden)

A dialogue manager for natural language interfaces

NL Interface 用の dialogue manager の作成に関する調査と実施例に関して, 発表があった. まず, NL dialogues について, 計算機との対話と人間との対話との差についての考察 (状況, コミュニケーションのチャネル, 主体の差) があった. 次に Database インタフェース用に, Wizard of Oz 実験により収集した 150 の対話例について説明があった. その結果, focus structure と dialogue structure の 2 種類の構造があり, これらを CFG rule (12-14 個) を用いた Dialogue Tree により扱う提案があった.

4 おわりに

参加者が 50 人程度と, 適切な規模であったので, PA-CLING が重視している「アットホームな」雰囲気の中で, 活発な議論があった. なお, 日本語や韓国語の細部に関する発表もあったが, 質問が多数出た. また, 大学の Book Shop にも, 日本語や韓国語の辞書が多数あった. これらの言語に関する興味の高さ感じられる.

最後に, closing で, 次回は, 1997 年 11 月に, 明星大学情報学部電子情報学科 (東京都青梅市) で開催されることが発表された. 実施は, 榊博史先生が担当される. 多数の参加を期待したい.

PACLING - II Program

WEDNESDAY, 19 April

9.15 Opening:

Professor Naoyuki OKADA, President, PACLING

Professor Peter SHEEHAN,
Pro-Vice-Chancellor (Research and
Graduate Studies, The University of
Queensland)

9.30-10.30 Keynote address: Charles FILLMORE
(University of California at Berkeley)
Dictionaries in the twenty-first century

11.00-12.30 Session A: Understanding and parsing
Kenzi IMAMURA and Yoshifumi OYAMA
(NTT Information and Communication Systems Laboratories)
Identifying user models of the sender and receiver in
a message domain

Ken BARKER and Stan SZPAKOWICZ (University of Ottawa)
Interactive semantic analysis of clause-level relationships

Yoshio ERIGUCHI and Tsuyoshi KITANI
(NTT Data Communications Systems Corporation)
Information extraction from Japanese text using
Tomita's generalized LR parser

Session B: Information processing and text
Hiromi BABA, Tsukasa NOMA and Naoyuki OKADA
(Kyushu Institute of Technology)
Visualizing temporal and spatial information in
natural language stories

2.0-3.00 Keynote address: M.A.K. HALLIDAY
(University of Sydney)
Computing meanings: Some reflections on past
experience and present prospects

3.30-5.30 Session A: Machine translation
Francis BOND, Kentaro OGURA and Satoru IKEHARA
(NTT Communication Science Laboratories)
Possessive pronouns as determiners in
Japanese-to-English machine translation

Geunbae LEE, Hanmin JUNG and Jong-Hyeok LEE
(Pohang University)

Bi-directional memory-based dialogue translation: the
KEMDT approach

Hiroaki KAGAMI, Kyoko KAI and Jun-ichi NAKAMURA
(Kyushu Institute of Technology)
and Sho YOSHIDA (Kyushu Institute of Design)
Interactive Korean to Japanese translation tool using
sequential selection method

Session B: Information processing and abstraction
Ichiro KOBAYASHI (Macquarie University) and
Michio SUGENO (Tokyo Institute of Technology)
An approach to social system simulation based on human
information processing

Mark LAUER (Microsoft Institute)
How much is enough? Data requirements for statistical NLP

Ian THOMAS and Ingrid ZUCKERMAN (Monash University)
and Bhavani RASKUTTI (Telecom Research Laboratories)
Accounting for pronunciation of phonemes in corpora

THURSDAY, 20 April

9.30-10.30 Keynote address: Michio SUGENO
(Tokyo Institute of Technology)
Intelligent fuzzy computing

11.00-12.30 Session A: Language Description for NLP
Shiho OGINO (IBM Research Laboratory, Tokyo)
An investigation of Japanese morphological
characteristics based on observation of the rules
applied in morphological analysis

Wei LI and Paul McFETRIDGE (Simon Fraser University)
Handling Chinese NP predicate in HPSG

One-Soon HER (National Chengchi University, Taipei)
Modularity and explanation: the case of VO
constructions in Chinese

Session B: Understanding and parsing
Sylvain DELISLE (Universite de Quebec)
and Stan SZPAKOWICZ (University of Ottawa)
Realistic parsing: practical solutions of difficult
problems

Hiroshi SAKAKI (Meisei University)

A syntactic to dependency structural transformation
method using piecewise tree transformation

Hyun S. PARK (University of Cambridge)
Handling of scrambling in Korean using MC-TAGS

2.00-3.30 Session A: Dictionaries and disambiguation
Fumiyo FUKUMOTO (Yamanashi University)
and Jun-ichi TSUJII (UMIST)
Word-sense disambiguation using the extracted
polysemous information from corpora

Alpha K. LUK (Microsoft, Sydney)
Sense disambiguation of frequently used polysemous nouns

Session B: Word tokens and identification
Kenji ARAKI, Yoshio MOMOUCHI (Hokkai-Gakuen University)
and Koji TOCHINAI (Hokkaido University)
Evaluation for adaptability of Kana-Kanji translation
of non-segmented Japanese Kana sentences using
inductive learning

Tetsuo ARAKI, Satoru IKEHIRA, Youichi SANADA
and Hiroshi YOSHINAGA (Fukui University)
An evaluation of a method to correctly select sentence
candidates of "Kanji-Kana" character strings using
Markov model

4.00-5.00 Session A: Dictionaries and disambiguation
Tony C. SMITH and Ian H. WITTEN (University of Waikato)
Probability-driven lexical classification: a
corpus-based approach

Peter WALLIS and James A. THOM (RMIT)
A better dictionary from LDOCE

Session B: Word tokens and identification
Masahiro OKU and Koji MATSUOKA (NTT Information and
Communication Systems Laboratories)
Detecting Japanese homophone errors in compound nouns
from character cooccurrence

Takehiro NAKAYAMA (Fuji Xerox Palo Alto Laboratory)
Text categorization using word shape tokens

FRIDAY, 21 April

10.00 Excursion (optional) to Currumbin Sanctuary, Gold
Coast
Evening Conference dinner

SATURDAY, 22 April

9.30-10.30 Keynote address: John BATEMAN
(Gesellschaft fuer Mathematik und Datenverarbeitung,
Germany)

Why get a translation when you can say it yourself? --
multilingual text generation and information systems

11.00-12.30 Session A: Generating, understanding
and parsing
Robert DALE (Microsoft, Sydney)
Generating one-anaphoric expressions: where does the
decision lie?

Richard F.E. SUTCLIFFE, Tom BREHONY and
Annette McELLIGOTT (University of Limerick)
The grammatical analysis of technical texts using a
Link Parser

Hongliang QIAO (The University of Queensland)
Extracting syntactic knowledge from a corpus for
Natural Language Parsing

Arlene HARVEY and Licheng ZENG (University of Sydney)
Patterns of definitions: towards tracking terms in
technical documents

Session B: Interfaces and learning
Arne JOHNSSON (Monash University)
A dialogue manager for natural language interfaces

Munindar P. SINGH and Mona SINGH (MCC)
A natural language interface to information retrieval
systems

Kyuko KAI and Jun-ichi NAKAMURA
(Kyushu Institute of Technology)
An interpreting and tutoring method for Japanese
interpersonal expressions

PACLING-II COMMITTEES

Organizing Committee

Chair:

Naoyuki Okada (Kyushu Institute of Technology, Japan)

Members:

Naoyuki Okada (Kyushu Institute of Technology, Japan)

Christian Matthiessen (University of Sydney, Australia)

Nick Cercone (Simon Fraser University, Canada)

Charles Fillmore (University of California, Berkeley, USA)