第十八屆自然語言與語音處理研討會
(ROCLING XVIII)

會議日期：95 年 9 月 6 日~9 月 8 日
會議地點：Tutorial-台北中研院活動中心第一會議室
Main Conference-交通大學電子與資訊中心
主辦單位：交通大學電信工程系、中華民國計算語言學學會(ACLCLP)

Tutorial Program

September 6		
Time	**Session**	**Speaker**
09:20-09:30	**Registration**	
09:30-10:40	Ontologies and NLP (1/2)	Laurent Prevot
10:40-11:00	**Coffee break**	
11:00-12:20	Ontologies and NLP (2/2)	Laurent Prevot
12:20-13:30	**Lunch break**	
13:30-14:30	New Resource in Chinese Language Processing - Fully Tagged Chinese GigaWord Corpus	Chu-Ren Huang & Keh-Jiann Chen
14:30-14:50	**Coffee break**	
14:50-16:50	Corpus Management and Processing Tools	Yuji Matsumoto

Tutorial Abstract- Laurent Prevot

Ontologies and Natural Language Processing

Laurent Prevot (Academia Sinica)

This talk provides/concerns the use of ontologies in computational linguistics. It will first define ontologies and explores their diversity from foundational ontologies to domain ones. Then it will detail their applications in computational linguistics but also the use of Natural Language Processing for building and improving them. A special attention will be given to linguistic ontologies such as famous lexical resources (WordNet and FrameNet) and their combination with traditional ontologies.

Tutorial Abstract - Chu-Ren Huang and Keh-Jiann Chen

New Resources in Chinese Language Processing - Fully Tagged Chinese GigaWord Corpus

Chu-Ren Huang and Keh-Jiann Chen
(Academia Sinica)

The Chinese GigaWord Corpus (CGW Corpus hereafter), first released by LDC in 2003 and updated in 2005, has the following crucial characteristics:

- CGW Corpus is the largest publicly available Chinese Corpus. CGW Corpus 1.0 contains more than 1,100 million characters, while CGW Corpus 2.0 contains nearly 1,300 million characters.
- CGW Corpus is the only corpus that containing sizable data from both PRC and Taiwan. CGW Corpus 2.0 contains more than 790 million characters from Taiwan, more than 470 million characters from PRC, as well as more than 28 million characters form Singapore.

The complete CGW Corpus is easy to access since internal formatting has been unified, and each text is clearly marked and classified by topic. However, neither versions 1.0 or 2.0 were segmented or tagged. In order to turn CWG Corpus as the basic resource for more versatile Chinese language processing in the future, a fully tagged version of the CGW Corpus was prepared (Ma and Huang 2006), and will be made available shortly. This tutorial introduces the tagged CGW Corpus.

The following topics will be covered in this tutorial:

- the content and composition of CGW Corpus, including its text formatting and topic classification
- the tagset adopted (CKIP-SinicaCorpus tagset)
- the tagging methodology
- quality assurance: specific improvements and independent evaluation
- availability of the data: Licensing from LDC and browsing Chinese WordSketch

Selected References

Chinese GigaWord Corpus
http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2003T09
Chinese GigaWord Corpus Second Edition
http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2005T14
Wei-yun Ma, and Chu-Ren Huang. 2006. Uniform and Effective Tagging of a Heterogeneous Giga-word Corpus. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC2006). Genoa, Italy. 24-28 May, 2006.

Tutorial Abstract - Yuji Matsumoto (NAIST)

Corpus Management and Processing Tools

Yuji Matsumoto (NAIST)

Annotated corpora are valuable resource for natural language processing as well as linguistic research. We have been developing a corpus management tools that store POS and syntactic dependency annotated corpora and provide various functions such as search and visualization.

This tutorial introduces the system together with freely available language processing tools we are developing, such POS taggers, chunkers and dependency structure analyzers.

Main Conference Program

September 7		
Time	**Session**	**Chair**
08:30-09:00	**Registration**	
09:00-09:10	Opening Ceremony-陳信宏教授	
09:10-10:10	Keynote Speech : Prof. Yuji Matsumoto	黃居仁教授
10:10-10:40	**Coffee Break**	
10:40-12:00	Session 1 – 語法及語意處理	高照明教授
12:00-13:20	**Lunch** **中華民國計算語言學會會員大會**	
13:20-14:20	Invited Speech – I : 簡立峰教授	陳信宏教授
14:20-14:40	**Break**	
14:40-15:40	Session 2 – 語音信號處理	廖元甫教授
15:40-16:00	**Coffee Break**	
16:00-16:50	Panel Discussion – Questions from Speech Technology Development to Linguistic and Phonetics – Dialogue and Future Directions 語音科技與語音學對談：談談語音科技開發過程中的語言語音問題	鄭秋豫博士
18:00-	**Banquet**	

September 8		
Time	**Session**	**Chair**

09:10-10:10	Invited Speech – II : 黃居仁教授	張俊盛教授
10:10-10:40	**Coffee Break**	
10:40-12:00	Session 3 – 語音及語者辨認	王新民教授
12:00-13:20	**Lunch**	
13:20-14:20	Session 4 – 資訊檢索，文件分類及語意網	盧文祥教授
14:20-14:40	**Break**	
14:40-16:00	Session 5 – 語言模型及其應用	張景新教授
16:00-16:20	**Closing**	

Session 1 語法及語意處理

1. 以語料為基礎的中文語篇連貫關係自動標記
   鄭守益、吳典松、梁婷
2. 中文動詞名物化判斷的統計式模型設計
   馬偉雲、黃居仁
3. 大規模詞彙語意關係自動標示之初步研究 - 以中文詞網為例
   Petr Šimon、謝舒凱、黃居仁
4. Automatic Learning of Context-Free Grammar
   Tai-Hung Chen、Chun-Han Tseng
5. Improve Parsing Performance by Self-Learning
   Yu-Ming Hsieh、Duen-Chi Yang、Keh-Jiann Chen

Session 2 語音信號處理

1. 國語雙字語詞聲調評分系統
   古鴻炎、孫世諺、張小芬
2. 一種用於網路電話之遺失封包補償方法
   古鴻炎、陳佳新
3. 基於字詞內容之適應性對話系統
   朱育德、張嘉惠
4. 一種適用於大量連續語料的語音文句校準方法
   簡世杰、張信常

Session 3 語音及語者辨認

1. 利用聲學與文脈分析於多語語音辨識單元之產生

王士豪、黃建霖、吳宗憲

2. 統計圖等化法於雜訊語音辨識之進一步研究
   林士翔、葉耀明、陳柏琳
3. 應用不定長度特徵之條件隨機域於口語不流暢語流修正
   葉瑞峰、吳維彥、吳宗憲
4. 鑑別性事前資訊應用於強健性語音辨識
   丁川偉、吳柏樹、簡仁宗
5. 結合韻律與聲學訊息之強健性漢語語者驗證系統
   張文杰、陳鼎允、陳子和、曾志仁、廖元甫、莊堯棠

Section 4 資訊檢索，文件分類及語意網

1. Personalized Optimal Search in Local Query Expansion
   Shan-Mu Lin、Chuen-Min Huang
2. 以本體論為基礎之新聞事件檢索與瀏覽
   許孟淵、黃純敏
3. 以部落格文本進行情緒分類之研究
   楊昌樺、陳信希
4. MiniJudge: Software for minimalist experimental syntax
   James Myers

Section 5 語言模型及其應用

1. 基於特製隱藏式馬可夫模型之中文斷詞研究
   林千翔、張嘉惠
2. 以字串特徵做為文本資料之錯誤偵測
   劉吉軒、鄭雍瑋
3. Learning to Parse Bilingual Sentences Using Bilingual Corpus and Monolingual CFG
   Chung-Chi Huang、Jason S. Chang
4. 使用流暢性改善詞組翻譯的統計式機器翻譯
   夏敏翔、張耀升、盧文祥
5. An Evaluation of Adopting Language Model as the Checker of Preposition Usage
   Shih-Hung Wu、Chen-Yu Su

September 7, 2006
ROCLING 2006

Machine Learning-based NLP Systems and Application to Opinion Mining

Yuji Matsumoto
Graduate School of Information Science
Nara Institute of Science and Technology
(NAIST, Japan)

Corpus-based NLP Research in our Group

NAIST

1. Natural Language Analysis Tools
   - POS taggers, Chunkers, Dependency Parsers of Japanese, Chinese and English
   - Application of Machine Learning for High performance NL analysis
2. Applications
   - Opinion Mining
   - Information Extraction (NE, Term, Evidence)
3. Management Tools for Linguistic Data
   - Annotated Corpus Management Tool: ChaKi
   - Dictionary Management Tool: Cradle

From the View of Corpus Annotation

NAIST

1. Natural Language Analysis Tools — Annotation
   - POS taggers, Chunkers, Dependency Parsers of Japanese, Chinese and English
   - Application of Machine Learning for High performance NL analysis
2. Applications — Use
   - Opinion Mining
   - Information Extraction (NE, Term, Evidence)
3. Management Tools for Linguistic Data
   - Annotated Corpus Management Tool: ChaKi
   - Dictionary Management Tool: Cradle — Management

Corpus Use: Opinion Mining (Mining Opinions with Evidence)

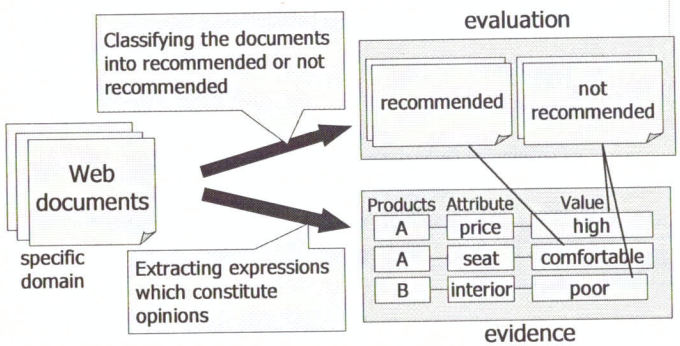
NAIST

It is important to know if people like or dislike things (opinion) as well as how (the reason)

- In consumer product survey:
  - People's opinion gives good indication of how to improve their products
  - Text mining of Weblog data or enquiry/survey texts helps this task
- In Evidence Based Medicine (EBM):
  - Text mining of medical documents/papers helps to find evidence of medical trials (of new drags/treatment)

NAIST

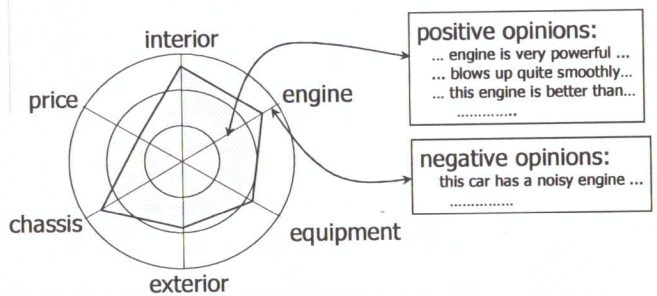
Our Approach to Opinion Mining



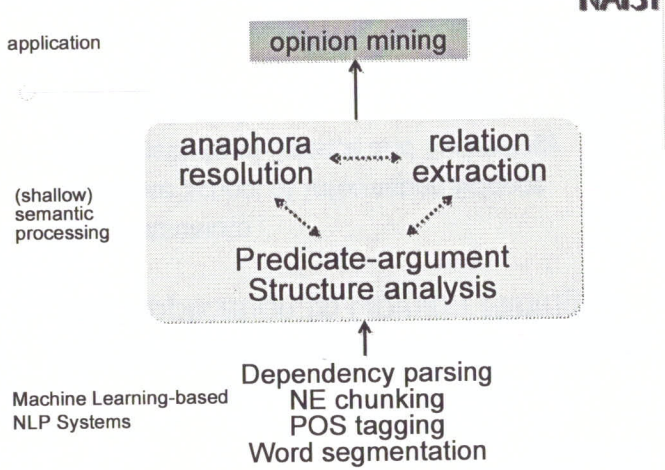
NAIST

Summarization of Extracted Opinions

◆Example: Radar chart representation



NAIST



NAIST

Basic Language Analyses (POS-tagging, phrase chunking, parsing)

Raw sentence

He reckons the current account deficit will narrow to only 1.8 billion in September.

Part-of-speech tagging

POS-tagged sentence

He reckons the current account deficit will narrow to only 1.8 billion in September .

PRP VBZ DT JJ NN NN MD VB TO RB CD CD IN NNP .

Base phrase chunking

Base pharase-chunked sentence

He reckons the current account deficit will narrow to only 1.8 billion in September .

NP VP NP VP PP NP PP NP

Dependency parsing

Dependency parsed sentence

He reckons the current account deficit will narrow to only 1.8 billion in September .

Basic Language Analyses (POS-tagging, word depedency parsing)

NAIST

Raw sentence

He reckons the current account deficit will narrow to only 1.8 billion in September.

Part-of-speech tagging

POS-tagged sentence

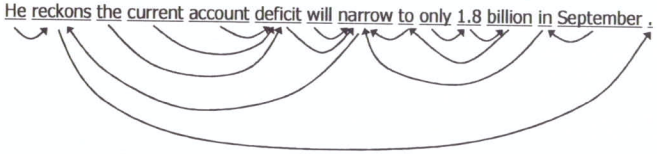
He reckons the current account deficit will narrow to only 1.8 billion in September.

PRP VBZ DT JJ NN NN MD VB TO RB CD CD IN NNP .

Word dependency parsing

Word dependency parsed sentence

He reckons the current account deficit will narrow to only 1.8 billion in September .



Base Phrase Chunking and Dependency Parsing of Japanese sentences

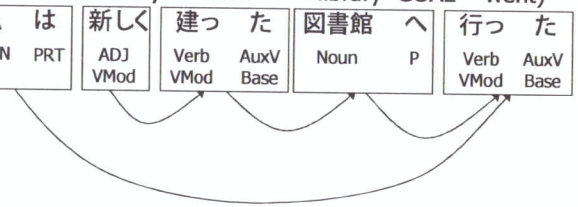
NAIST

Dependency analysis

Base phrase chunking

(I 1TOPIC newly built library GOAL went)

私 は	新しく	建っ た	図書館 へ	行っ た
PRN PRT	ADJ VMod	Verb AuxV VMod Base	Noun P	Verb AuxV VMod Base



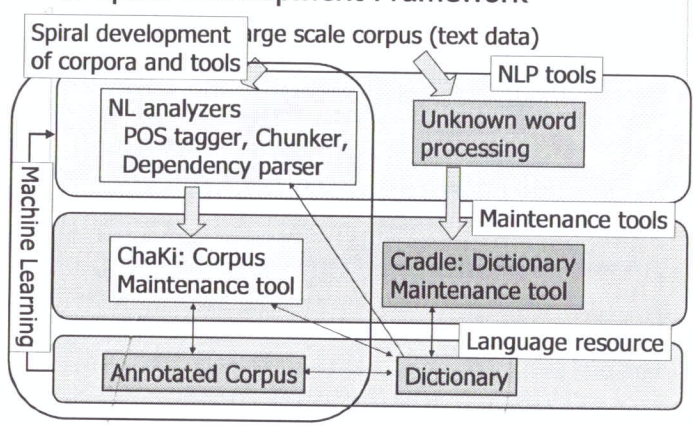
Corpus-based research activities conducted in our Group

NAIST

- NLP Tools Based on Machine Learnir — Corpus Annotation
  - Japanese Morphological Analyser: ChaSen [Asahara 00] [Matsumoto 00, 04]
    - Multi-lingual version: Japanese, Chinese, English
  - Japanese Dependency Parser: CaboCha [Kudo 02]
  - English and Chinese Word Dependency Parsers [Yamada03,Chen04]
  - General Purpose Chunker: YamCha [Kudo 01]
    - Named Entity Recognition [Asahara 03]
    - Unknown Word Identifier: bar [Asahara 04]
  - Anaphora Resolution
    - Finding antecedent of Japanese zero-pron
- Management Tools for Linguistic Data — Annotated Corpus Management
  - Annotated Corpus Management System: ChaKi [Matsumoto 05]
  - Dictionary Management System: Cradle

NLP and Corpus Maintenance Tools in Spiral Development Framework

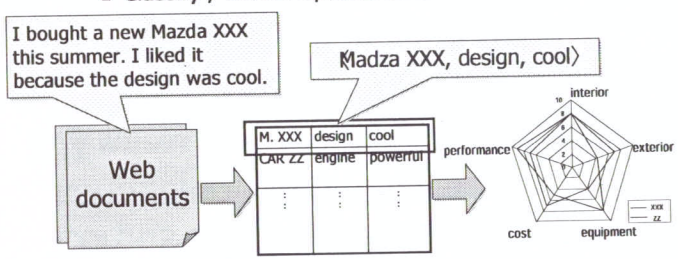
NAIST



NAIST

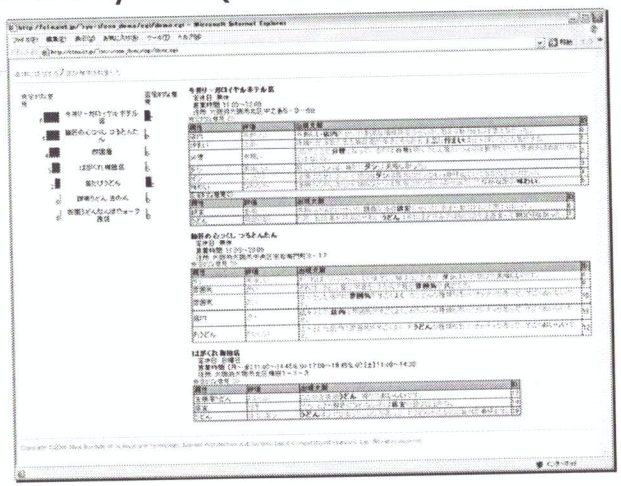
Opinion mining

- Aim: Analyzing opinions that describe how customers feel about certain aspects of certain subjects (products, shops, services, etc.)
- Tasks:
  - Extract opinion units ⟨subject, aspect, evaluation⟩
  - Classify / cluster opinion units



NAIST

Demo system (restaurant reviews)



NAIST

Demo system (restaurant reviews)

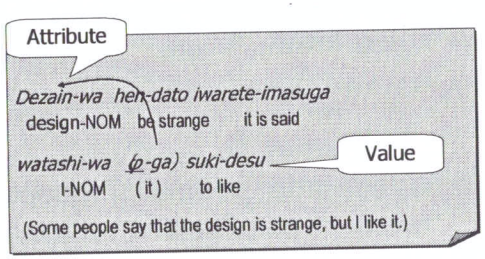
はがくれ梅田店
定休日: 日曜日
営業時間: [月～金] 11:00～14:45(L.O) 17:00～19:45(L.O) [土] 11:00～14:30
住所: 大阪府大阪市北区梅田１－１－３
肯定的な意見 (24)

属性	評価	出現文脈	ID
食感	旨い	「うどんは調身」と店主が言っているだけあって、しっかりしたコシ、ツルツルした**食感**でとても旨い。	0
コシ	しっかり	「うどんは調身」と店主が言っているだけあって、しっかりした**コシ**、ツルツルした食感でとても旨い。	1
生醤油うどん	おいしい	ここの生醤油**うどん**、確かにおいしいです。	2
店主	好き	でも、どうも好きになれないのは**店主**の語り口上かな。	3
	最高	**混んでいなければ最高のお店です。**	4
うどん	おいしい	ぶっかけ**うどん**は、やっぱおいしいです。	5
うどん	満足	ぶっかけ**うどん**は、やっぱおいしいです。本場の讃岐に行かないと食べられないのかなと思っていたが、ここなら満足。	6
雰囲気	最高	お店の**雰囲気**は最高ですが、混んでいなければね。	7
知名度	高い	**知名度**においては、「今井」と並び、最も高い一軒でしょう。	8
回転	良い	ただし、昼の**回転**は比較的良く、	9
生醤油うどん	勧める	是非とも名物の「生醤油**うどん**」を御勧めしたい。	10
感じ	良い	私は、むしろ**感じ**の良い人に思います。	11

NAIST

Opinion mining

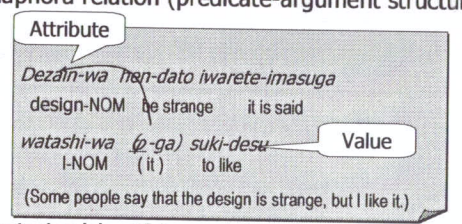
- Extracting opinions by relation extraction and zero-anaphora relation (predicate-argument structure)



NAIST

Opinion mining

- Extracting opinions by relation extraction and zero-anaphora relation (predicate-argument structure)



- Clustering/classifying opinions by paraphrase/entailment recognition
  - heavy ⇔ weighty
  - I like … ⇔ I'm satisfied with …
  - has a big trunk ⇔ has enough storage space ⇔ spacious

NAIST

Noun phrase anaphora resolution

- Anaphora resolution is the process of determining whether two expressions in natural language refer to the same real world entity

A federal judge in Pittsburgh issued a temporary restraining order preventing Trans World Airlines from buying additional shares of USAir Group Inc. antecedent anaphor
The order, requested in a suit filed by USAir, dealt another blow to TWA's bid to buy the company for $52 a share.

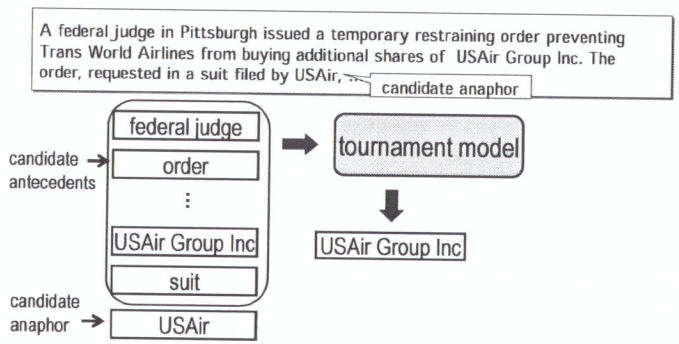
- An anaphor may be a noun phrase, a pronoun, or missing (e.g., Japanese zero pronouns)
- Important process for various NLP applications: machine translation, information extraction, question answering

NAIST

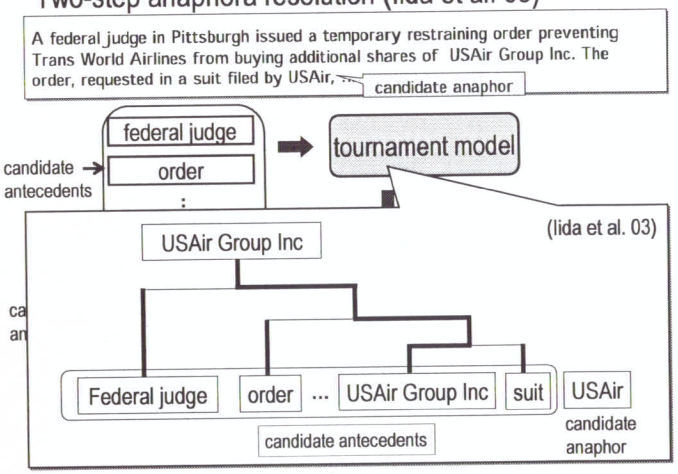
Background

- Rule-based approach
  (Mitkov 97, Baldwin 95, Nakaiwa et al. 96, etc.)
- Learning-based approach
  - Recasting anaphora resolution as classification problems (Soon et al. 01, Ng and Cardie 02, Iida et al. 2003, etc.)
  - Comparable to the state-of-the-art rule-based system
  - Anaphoricity determination increasingly paid attention (Bean and Riloff 99, Uryupina 03, Ng 04, Iida et al. 05, etc.)

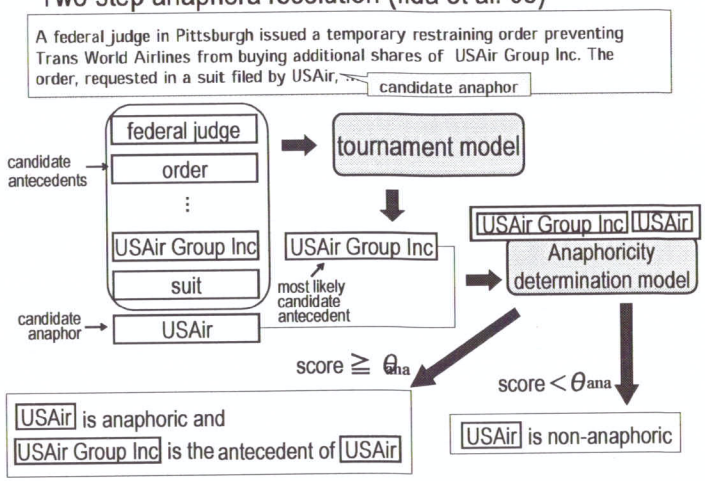
Two-step anaphora resolution (Iida et al. 05)



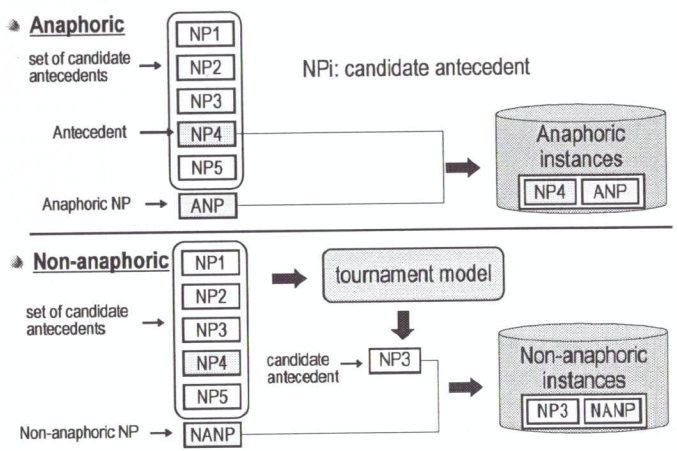
Two-step anaphora resolution (Iida et al. 05)



Two-step anaphora resolution (Iida et al. 05)



Training the anaphoricity determination model



Experiments

- NP anaphora resolution in Japanese
- Japanese newspaper article corpus annotated with NP-anaphoric relations
  - 90 text (1,104 sentences)
  - Noun phrases : 876 anaphors and 6,292 non-anaphors

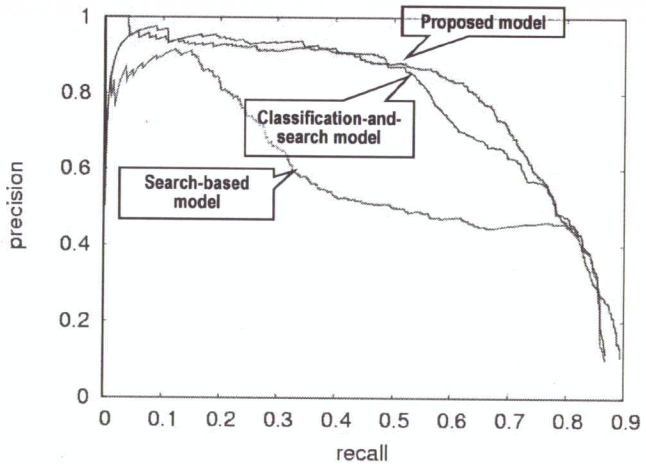
$$\text{Recall} = \frac{\text{\# of correctly detected anaphoric relations}}{\text{\# of anaphoric NPs}}$$

$$\text{Precision} = \frac{\text{\# of correctly detected anaphoric relations}}{\text{\# of NPs classified as anaphoric}}$$

Comparison with previous approaches

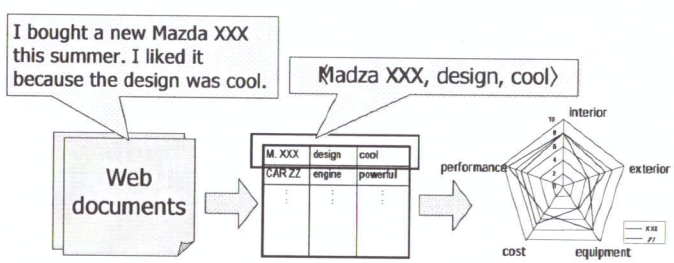
- Search-based approach (SM)
  (Soon et al. 01, Ng and Cardie 02)
  - Recast anaphora resolution as binary classification problems
  - Comparable to the state-of-the-art rule-based system
  - Disadvantage: not use non-anaphoric instances in training
- Classification-and-search approach (CSM)
  (Ng and Cardie 02, Ng 04)
  - Determine anaphoricity before identifying antecedents
  - The performance of the CSM is better than the SM if the threshold parameters are appropriately tuned
  - Disadvantage: not use the contextual information (i.e. whether an appropriate antecedent appears on the context)

NAIST



Opinion mining

- Aim:
  - Analyzing opinions that describe how customers feel about certain aspects of certain subjects (products, shops, services, etc.)
- Tasks:
  - Extract opinion units ⟨subject, aspect, evaluation⟩
  - Classify / cluster opinion units



Structuring opinions

I bought a powershot last week from amazon.
I took hundreds of pictures.
great colors and white balance even when flash is used.

opinion units

opinion holder	⟨writer⟩	opinion holder	⟨writer⟩
subject	⟨powershot ⟩	subject	⟨powershot ⟩
part	⟨picture⟩	part	⟨picture⟩
attribute	⟨colors⟩	attribute	⟨white balance⟩
evaluation	⟨great⟩	evaluation	⟨great⟩
condition	⟨flash is used⟩	condition	⟨flash is used⟩

Corpus study

	Restaurant	Automobile	Cellphone	Game
# of texts	1,356	564	481	361
# of sents	21,666	14,005	11,638	6,448
Asp-Eval	3,692	943	965	521
Asp-Asp	1,426	280	296	221
Subj-Asp	2,632	877	850	451
Support	68	86	80	95
Counter part	32	66	75	41
Condition	113	86	76	41

Corpus study

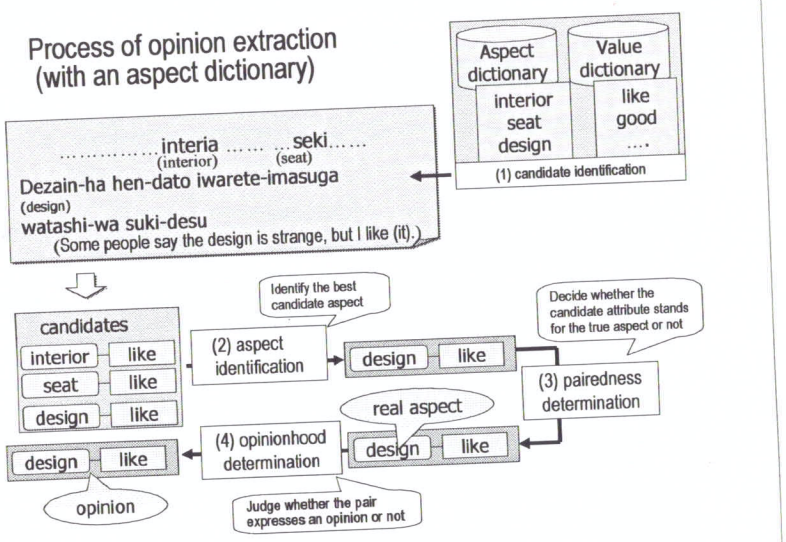
	Restaurant	Automobile	Cellphone	Game
# of opinion units	4,267	1,519	1,518	775
Subj-Eval	575	576	553	243
Subj-Asp-Eval	2,314	736	768	351
Subj-Asp-Asp-Eval	1,065	175	172	127
Subj-Asp-…-Eval	313	32	25	54

Opinion extraction: Task

	Restaurant	Automobile	Cellphone	Game
# of opinion units	4,267	1,519	1,518	775
Subj-Eval	575	576	553	243
Subj-Asp-Eval	2,314	736	768	351
Subj-Asp-Asp-Eval	1,065	175	172	127
Subj-Asp-…-Eval	313	32	25	54

- Extraction of
  〈opinion-holder, subject, part, attribute, evaluation〉

Process of opinion extraction (with an aspect dictionary)



- Aspect-evaluation relation extraction

I bough a powershot last week from amazon. Today's weather is fine, so I went to the park to take some pictures. I took hundreds of pictures. Great colors even when flash is used.

Opinion-hood determination

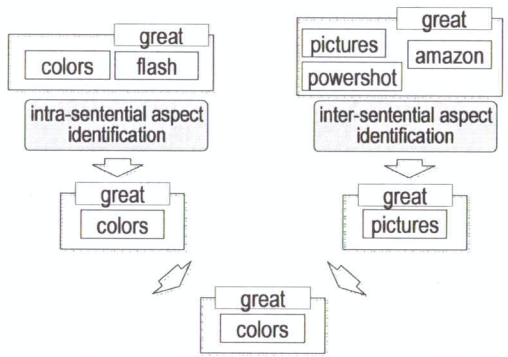
I bough a powershot last week from amazon. Today's weather is fine, so I went to the park to take some pictures. I took hundreds of pictures. Great colors even when flash is used.

- Aspect-of relation extraction

I bough a powershot last week from amazon. Today's weather is fine, so I went to the park to take some pictures. I took hundreds of pictures. Great colors even when flash is used.

Aspect-evaluation relation extraction

I bough a powershot last week from amazon. I took hundreds of pictures and they were excellent. Great colors even when flash is used.



"Aspecto-hood"

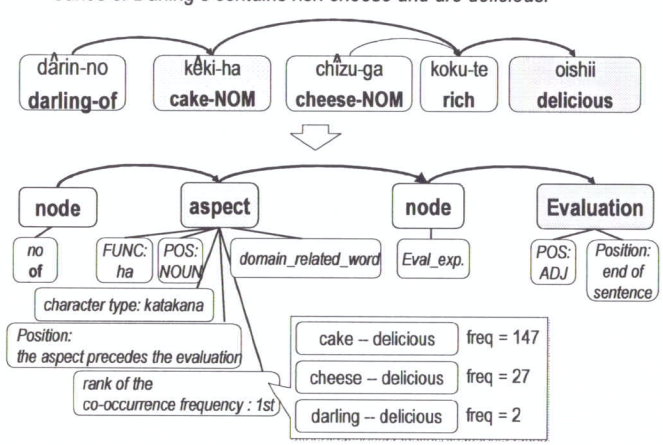
- Coocurrence with the target domain / domain subjects

Coocurrence frequency

taste (365,120), flavor (59,110), quantity (37,400)
meat (24,061), soup (5,187), color (4,777)
cream (3,087), noodle (2,938), master (2,833)
cake (1,863), cheese (1,645), set meal (1,602)
menu item (1,523)

Intra-sentential aspect identification model

Cakes of Darling's contains rich cheese and are delicious.



Results of aspect-evaluation relation extraction

		intra-sentential	inter-sentential	total
baseline	precision	0.44 (250/570)		0.44 (250/570)
	recall	0.31 (250/809)		0.23 (250/1083)
proposed model	precision	0.61 (573/809)	0.30 (44/146)	0.57 (617/1083)
	recall	0.71 (573/809)	0.16 (44/274)	0.57 (617/1083)
proposed model + dictionary	precision	0.69 (601/870)	0.35 (74/212)	0.62 (675/1083)
	recall	0.74 (601/809)	0.27 (74/274)	0.62 (675/1083)

Results of aspect-of relation extraction

		intra-sentential	inter-sentential	total
baseline	precision	0.24 (44/180)		0.24 (44/180)
	recall	0.44 (44/101)		0.16 (44/270)
proposed model	precision	0.51 (45/89)	0.21 (38/181)	0.31 (83/270)
	recall	0.45 (45/101)	0.23 (38/169)	0.31 (83/270)
proposed model + dictionary	precision	0.59 (48/81)	0.24 (45/185)	0.34 (93/270)
	recall	0.48 (48/101)	0.26 (45/169)	0.34 (93/270)

NAIST

On-going work

- Classification of opinion units
  - Usability, functionality, cost, design, etc.
- Collaboration with companies
  - 150 million blog posts
  - Marketing
    - Mining customer reviews
    - Tracking behaviors of (potential) customers

NAIST

Summary

- Corpus-based (machine learning (ML)-based) Natural Language Processing
  - ML-based NLP tools: POS-tagging, Chunking, Dependency Parsing, Anaphora resolution
  - For Accurately Annotated Corpus Creation
  - Annotated corpus management tools
- Application of Natural Language Processing
  - Opinion mining with evidence

Natural Language Processing versus Web Search Engines

Lee-Feng Chien
Google Taiwan & Academia Sinica

A Web search engine is a computer system that helps users of the Internet locate information on the World Wide Web. It collects and indexes a variety of Internet resources and users' data. The greatest success in Web search engines has accumulated a huge amount of corpus sources in various domains, and acquired a huge amount of online users as testers for sophisticated software. Search engines have become a very well test bed for development of more advanced natural language technologies, such as question answering, information extraction, machine translation, cross-language information retrieval and spoken information retrieval, etc. In this talk, I would like to briefly introduce the infrastructure of a Web search engine and discuss a couple of NLP techniques that can benefit from the development of Web search engines.

Toward a Global Wordnet Grid: Infrastructure for Language Technology in the Age of Multilingualism

全球詞網網格倡議：多語社會中的語言科技基礎建設

黃居仁
中央研究院語言學研究所
churen@gate.sinica.edu.tw
http://cwn.ling.sinica.edu.tw/huang/huang.htm

Multilingualism is potentially the most rewarding challenge that language technology will face in the near future. A critical part of the challenge is the scaling up of language resources in a complex and distributed environment. Language resources, lexicons included, are inherently distributed because of the diversity of language distributed over the world. It is natural and efficient for language resources to be developed and maintained in their native environment. In addition, since language evolves and changes over time, it is not possible to describe the current state of the language away from where the language is spoken. Lastly, the vast range of diversity of languages also makes it impossible to have one single universal centered resource.

An international initiative has been discussed recently to solve this challenge. The initiative is two-prolonged: the first is to take wordnet as the shared format of linguistic resource to ensure inter-operability as well as to encode rich linguistic knowledge. The second is to borrow the infrastructure of grid computing in order to best utilize the distributed nature of language resources. In other words, we aim to link all each resource points to a grid, and to enable each resource point as a node in a computing grid.

Technical details that need to be solved in this initiative include, but are not limited to the following: proposing a standard format for wordnet encoding, identifying s common shared core lexicon for all languages, develop web service tools that will allow remote access and manipulation of multilingual resources. The first two issues are partially addressed in the first preliminary call for participation (http://www.globalwordnet.org/gwa/gwa_grid.htm), while all the above issues were discussed in a recent paper (Soria et al. 2006).

The fact that lexical resources are distributed over the world in essence makes their global distribution grid-like. This was a important motivation behind the global wordnet grid initiative. However, does the grid-like distribution of language resources lend itself to grid-computing? We think so,

especially in terms of multilingual processing. Any multilingual process, such as cross-lingual information retrieval, must involve both resources and tools in a specific language and language pairs. For instance, a multilingual query given in Italian but intended for querying English, Chinese, French, German, and Russian texts, can be send to five different nodes on the Grid for query expansion, as well as performing the query itself. In this way, language specific query techniques can be applied in parallel to achieve best results that can be integrated in the future.

Selected Bibliography

The Global Wordnet Association. http://www.globalwordnet.org/

Huang, Chu-Ren, Wan-Ying Lin, Jia-fei Hong, and I-Li Su. 2006. The Nature of Cross-lingual Lexical Semantic Relations: A Preliminary Study Based on English-Chinese Translation Equivalents. Proceedings of the Third International WordNet Conference. Pp. 180-189. Jeju. Januaray 22-25.

Soria, Claudia, Maurizio Tesconi, Andrea Marchetti, Francesca Bertagna, Monica Monachini, Chu-Ren Huang, and Nicoletta Calzolari. 2006. Towards Agent-based Cross-lingual Interoperability of Distributed Lexical Resources. Proceedings of the 2006 COLING/ACL post-conference workshop 'Multilingual Language Resources and Interoperability.' July 23. Sydney, Australia.