

Chinese Temporal Process and Machine Learning-based Dependency Analyzer

4th COE Postdoctoral and Doctoral Researchers Technical Presentation

Yuchang CHENG

Computational Linguistics Laboratory
Graduate School of Information Science

Nara Institute of Science and Technology

July 19th, 2005

The research plan of COE

- ▶ Title: Using the results of dependency analysis for extracting and processing temporal information in Chinese
- ▶ Recognize the temporal meaning of event/occurrence and recognize the relations between events
- ▶ Applications:
 - Q&A, planning, temporal information database, machine translation.....

The research plan of COE

- ▶ Recognize the temporal meaning of event/occurrence and recognize the relations between events

NEWSPAPERS
2005 / 05 / 29 (星期日)
馬英九於前晚九時許**抵達**
苗栗，.....
.....**傅縣長表示**會予協助
競選國民黨主席.....
.....**連戰前往**中國訪問後
總統**宴請**馬英九....

OCCUR(馬英九**抵達**)
PAST(前晚九時許)

OCCUR(傅縣長**表示**)
PAST(report time)

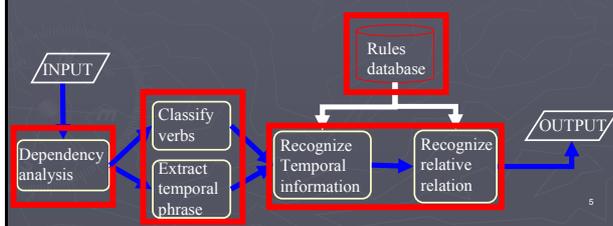
BEFORE((連戰**前往**),
(總統**宴請**))

Temporal information of Chinese

- ▶ Verbs don't change the morpheme according to tense
 - Cannot get temporal information directly from morpheme
- ▶ Temporal information
 - Verb class (temporal feature)
 - Temporal phrase
 - ▶ 在第二次世界大戰期間(During the WW II...)
 - Temporal adverbs
 - ▶ 在 (being), 已經 (have), 將要 (will)
 - Aspect auxiliary words
 - ▶ 了 (been), 過 (was)

Temporal Information Process System

- ▶ Using the results of dependency analysis for extracting and processing temporal information in Chinese



Temporal Information Process System (sub-systems)

- ▶ **Dependency analysis**
 - Related research (Li, 2001~2004) concluded that the temporal process needs syntactic analysis
 - The relation between verbs and temporal phrases are ambiguous
 - To perfectly cover the relations between verbs and temporal phrases needs complex rules data-base
- ▶ Extract temporal phrase and classify verb
- ▶ Define the rules
- ▶ Recognize Temporal information and the relative relation between events

Why Dependency Analyzer?

- ▶ The dependency structure can clearly explain the head-modifier relations between verbs and temporal phrases
- ▶ Using few and simple rules can extract the relations

Ex: 九六年飞弹试射的成功已改变两岸战略思维
 (The success of missile test in 1996 caused a change in the China and Taiwan's strategies.)

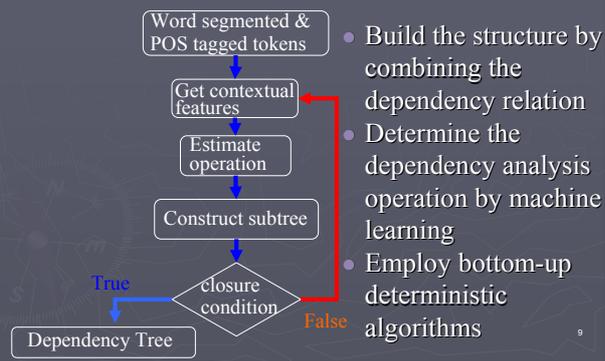


Note: The arrow direction is from modifier to head word

Chinese Dependency Analyzer

- ▶ Adopt deterministic parsing algorithms
 - Improve the [Nivre, 2004] algorithms
- ▶ Utilize Support Vector Machines to determine the word dependency relations
- ▶ Supply the phrase and root information to improve the analyzer
 - Extract Root word
 - Extract prepositional phrase

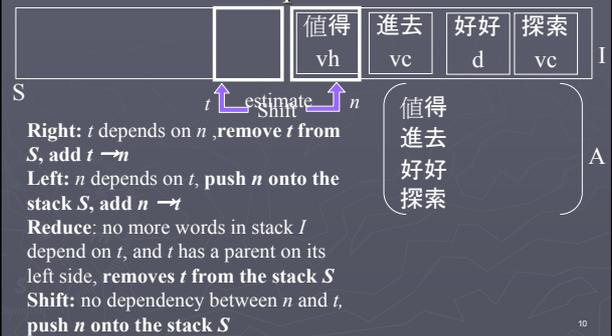
Procedure of Dependency Analysis



Algorithms

(original proposed by Nivre, 2004)

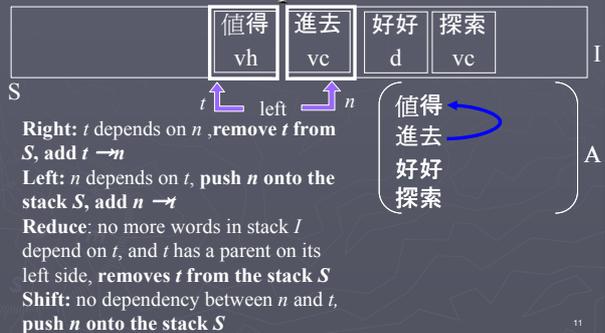
The operations



Algorithms

(original proposed by Nivre, 2004)

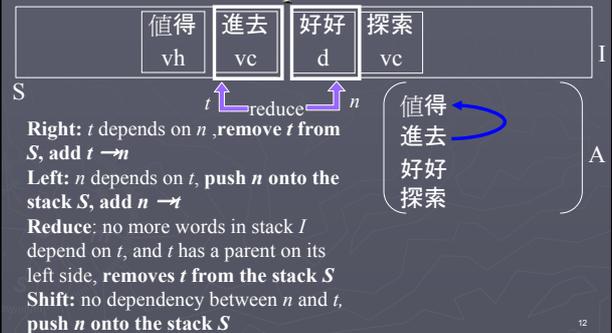
The operations



Algorithms

(original proposed by Nivre, 2004)

The operations



Algorithms

(original proposed by Nivre, 2004)

The operations

Right: t depends on n , remove t from S , add $t \rightarrow n$

Left: n depends on t , push n onto the stack S , add $n \rightarrow t$

Reduce: no more words in stack I depend on t , and t has a parent on its left side, removes t from the stack S

Shift: no dependency between n and t , push n onto the stack S

13

Algorithms

(original proposed by Nivre, 2004)

The operations

Right: t depends on n , remove t from S , add $t \rightarrow n$

Left: n depends on t , push n onto the stack S , add $n \rightarrow t$

Reduce: no more words in stack I depend on t , and t has a parent on its left side, removes t from the stack S

Shift: no dependency between n and t , push n onto the stack S

14

Improvement

adopt the root node finder

(To Leave native country to study and to visit other country.)

- ▶ Bottom-up analyzer
 - Lacks the top-down information
- ▶ The dependency relation cannot cross the root node
- ▶ propose : to divide the sentence by root node
 - Analyze each sub-tree independently
 - Construct a root finder based on SVMs
 - Precision : 90.63%

15

Improvement

Extract the prepositional phrase

(until the government correspond this problem...)

- ▶ The analyzer usually makes mistakes in analyzing prepositional phrase
 - Analyzer tends to analyze a preposition as governing a shorter rather than a longer phrase
- ▶ Need to extract the PP before dependency analysis
 - Analyzes the prepositional phrase first
 - Using a SVMs-based prepositional phrase chunker (YamCha)
 - Precision : 88.74%

16

Experimental Setting

- ▶ Corpus (Penn Chinese Treebank 5.0)
 - 377,406 words (45775 sentences) for training; 63,886 words (7030 sentences) for testing
- ▶ Utilize Support Vector Machines (SVM)
 - Pairwise, polynomial kernel (d=2)
- ▶ Features
 - node t and n , 2 preceding nodes of t , 2 succeeding nodes of n , and their child nodes.

17

	Dependency Accuracy(%)	Root Accuracy(%)	Sentence Accuracy(%)
Algorithm+SVMs	85.26	86.34	60.00
Algorithm+SVMs +root node	85.91	90.91	60.84
Algorithm+SVMs +PP chunker	85.75	87.07	60.93
Algorithm+SVMs +root node+PP chunker	86.11	90.91	61.79

18

Future directions

The research plan of COE (sub-plans)

- ▶ Dependency analysis (The Topic Today)
- ▶ Extract temporal phrase and classify verb
 - Recognize the meaning of various temporal phrase
 - ▶ Ex. 1950s=1950年代=約五十年前=二次大戰後10年
 - Classify verb
 - ▶ Meaning of verbs indicates the event type
 - ▶ Identify the event type by classifying verbs
- ▶ Define the rules
 - The relative relation between events
- ▶ Recognize Temporal information and the relative relation between events

19