

ROCLING XII - Program Schedule

8 月 26 日(星期四)

9:00-9:30	Registration	交通大學光復校區 電子資訊大樓
9:30-10:30	Opening and Keynote Speech	電子資訊大樓國際會議廳
10:30-11:00	Break	
11:00-12:05	Session 1: Text-to-Speech	電子資訊大樓國際會議廳
12:05-14:00	Lunch	
14:00-15:20	Session 2: Syntactic and Semantic Analysis	電子資訊大樓國際會議廳
15:20-15:50	Break	
15:50-16:30	Sessions 3: Text Categorization	電子資訊大樓國際會議廳
17:30-19:30	Banquet	科學園區天外天餐廳

8 月 27 日(星期五)

9:00-11:00	Tutorial	電子資訊大樓國際會議廳
11:00-11:20	Break	
11:20-12:20	ROCLING 會員大會	電子資訊大樓國際會議廳
12:20-13:40	Lunch	
13:40-15:00	Session 4: Information Retrieval and Extraction	電子資訊大樓國際會議廳
15:00-15:30	Break	
15:30-16:20	Sessions 5: Speech Recognition	電子資訊大樓國際會議廳

8 月 26 日(星期四)

9:30 ~ 10:30

Opening and Keynote Speech

Session Chair: 李琳山教授/台灣大學

From sentence structure to discourse connectedness 1

Speaker: Dr. Eva Hajicová/ Charles University, Prague, Czech Republic

11:00 ~ 12:05

Session 1 : Text-to-Speech

Session Chair: 簡仁宗教授/成功大學

臺語多聲調音節合成單元資料庫暨文字轉語音雛形系統之發展 15

余永吉、鍾高基、吳宗憲

國語文句翻台語語音系統之研究 37

王駿發、黃保章、林順傑

**Semantic classification of Patterns Containing Non-Text Symbols in
Mandarin Text** 55

Feng-Long Hwang, Ming-Shing Yu, Ming-Jer Wu, Shyh-Yang Hwang

14:00 ~ 15:20

Session 2: Syntactic and Semantic Analysis

Session Chair: 高照明教授/暨南大學

動詞詞構與語法功能互動初探 67

張麗麗、陳克健

**Semantic Representation of Verbal Information -- A Case from Mandarin
Verbs of Judging** 87

Mei-Chun Liu, Chu-Ren Huang, Jia-Ying Lee

An Analytical Study of Transformational Tagging on Chinese Text* 101

Helen M. Meng and Chun Wah Ip

On Modeling Remote and Local Dependencies in Language* 123

Yu-Sheng Lai and Chung-Hsien Wu

* short paper

15:50 ~ 16:30

Sessions 3: Text Categorization

Session Chair: 陳光華教授/台灣大學

階層式文件自動分類之特徵選取研究 137

柯淑津、陳振南

Term Selection with Distributional Clustering for Chinese Text

Categorization using N-grams* 151

Jyh-Jong Tsay and Jing-Doo Wang

8 月 27 日(星期五)

9:00 ~ 11:00

Tutorial

Session Chair: 蘇克毅教授/致遠科技公司

Dependency syntax in a formal framework 3

Speaker: Dr. Eva Hajicová/ Charles University, Prague, Czech Republic

13:40 ~ 15:00

Session 4: Information Retrieval and Extraction

Session Chair: 柯淑津教授/東吳大學

Automatically Controlled-Vocabulary Indexing for Text Retrieval 171

Kuang-hua Chen and Chien-tin Wu

A New Syllable-based Approach for Retrieving Mandarin Spoken

Documents Using Short Speech Queries 187

Hsin-min Wang

Recent Results on Domain-Specific Term Extraction From Online Chinese

Text Resources 203

Lee-Feng Chien, Chun-Liang Chen, Wen-Hsiang Lu, Yuan-Lu Chang

15.30 ~ 16:20

Sessions 5: Speech Recognition

Session Chair: 古鴻炎教授/台灣科技大學

國語電話語音辨認之強健性特徵參數及其調整方法 219

黃儀芬、王小川

音框同步之雜訊補償方法在汽車語音辨識之應用 239

簡仁宗、林敏順

* short paper

From Sentence Structure to Discourse Connectedness

Eva Hajicová

Charles University, Prague, Czech Republic

The Praguian notion of the information structure of the sentence (functional sentence perspective, topic-focus articulation, henceforth TFA) as analyzed in the writings of V. Mathesius, J. Firbas, F. Daneš, and, within an explicit theoretical framework, in Sgall et al. (1986) has found its counterparts in most different present-day theoretical frameworks and is viewed as an important aspect of the analysis of discourse, contributing to discourse cohesion and to the study of the use of anaphoric pronouns.

The dichotomy of topic and focus may be understood as the linguistic patterning basically corresponding to the cognitive opposition between 'given' ('known') and 'new' information, with the caveat that a 'given' item can be referred to as not 'known', esp. if put into contrast. A weak anaphoric pronoun can only be used as contextually bound, i.e. in the topic (unless embedded within the focus); in the prototypical case, it is coreferential with an item occurring in one of the closely preceding utterances, and it always refers to an entity that in the given time-point is above a certain threshold of salience in the stock of shared knowledge (SSK) assumed by the speaker to belong to the hearer's short-term memory; the original model we refer to in our paper was presented first by Hajicová and Vrbová in 1981, and is further developed in Hajicová 1987, Hajicová 1993, and Hajicová 1997.

SSK has a dynamic character: the repertoire of elements included in it as well as their degrees of activation (salience) change as the discourse progresses. The difference in the degrees of activation reflects the contention that some elements are relatively easier to access than some others. The decisions on a possibility/impossibility of pronominal reference on one side, and the resolution of anaphora on the other side thus can be based (among other issues, of course) on the difference of the degrees of activation.

An element is considered to be activated to a higher or lower degree depending on various factors. In our paper, based on an analysis of discourses of a monologue character, we will discuss some heuristics for the determination of this degree based basically on the topic-focus articulation of sentences; the degrees of activation impose a partial ordering on elements in the stock of shared knowledge, and it will be illustrated that an analysis in this way can also throw an interesting light on the segmentation of the text and on the notion of the 'topic' of the text.

References:

- Grosz Barbara J., Joshi Aravind K. and Scott Weinstein (1995), Centering: A Framework for modeling the local coherence of discourse. *Computational Linguistics* 21, No. 1, 203-225
- Hajicová Eva (1987), Focussing: a meeting point of linguistics and Artificial Intelligence. In: *Artificial Intelligence II - methodology, systems, application*, ed. By Ph. Jorrand and V. Sgurev. Amsterdam:North Holland, 311-322.
- Hajicová Eva (1993), *Issues of sentence structure and discourse patterns*. Prague: Charles University.

Hajicová Eva (1997), Topic, focus and anaphora In: Caron, Bernard, Ed., Actes du 16e Congres International des Linguistes (Paris 20-25 juillet 1997), Oxford (Elsevier Sciences).

Sgall Petr, Hajicová Eva and Jarmila Panevová (1986), The meaning of the sentence in its semantic and pragmatic aspects. Prague: Academia and Dordrecht: Reidel.