Table of Contents

Keynote Speech

Setting Parameters: Triggering and Mis-triggering 1
Janet D. Fodor

Invited Speech

Encoding Events across Languages 2
Beth Levin

Opposing Effects of Word and Character Frequency in the Processing of Chinese text 3
Ovid T.L. Tzeng, Chih Wei Hue, Daisy L. Hung

Session 1a: Lexical Semantics

Meaning Representation and Meaning Instantiation for Chinese Nominals 4
Kathleen Ahrens, Lily Chang, Keh-Jiann Chen, Chu-Ren Huang

Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy 19
Jay J. Jiang, David W. Conrath

Towards a Representation of Verbal Semantics - An Approach Based on Near-Synonyms 34
Mei-Chih Tsai, Chu-Ren Huang, Keh-Jiann Chen, Kathleen Ahrens

*Word Sense Disambiguation Based on the Information Theory 49
Ho Lee, Dae-Ho Baek, Hae-Chang Rim

Session 1b: Parsing

An Agreement Error Correction Method Based on a Multicriteria Approach: An Application to Arabic Language 59
Lamia Belguith Hadrich, Abdelmajid Ben Hamadou, Chafik Aloulou

Incorporating Bigram Constraints into an LR Table 76
Hiroki Imai, Hui Li, Hozumi Tanaka

A Level-Synchronous Approach to Ill-Formed Sentence Parsing 89
Yi-Chung Lin, Keh-Yih Su

*The Application of the Similarities between the Morphemes of the English and Chinese Languages to Represent Chinese Characters Phonetically with English Letters to Facilitate Computer Applications Manually and by Voice with the Character-Based Languages Chinese, Japanese and Korean 109
Stanley K. Chan

Session 2a: Lexical Analysis

A Multivariate Gaussian Mixture Model for Automatic Compound Word Extraction 123
Jing-Shin Chang, Keh-Yih Su

Proper Name Extraction from Web Pages for Finding People in Internet 143
Hsin-Hsi Chen, Guo-Wei Bian

Unknown Word Detection for Chinese by a Corpus-based Learning Method 159
Keh-Jiann Chen, Ming-Hong Bai

Session 2b: Message Understanding

Analyzing the Complexity of A Domain with Respect to An Information Extraction Task 175
Amit Bagga

Human Judgment as a Basis for Evaluation of Discourse-Connective-based Full-text Abstraction in Chinese 195
Benjamin K T'sou, Hing-Lung Lin, Tom B.Y. Lai

An Assessment on Character-based Chinese News Filtering Using Latent Semantic Indexing 209
Shih-Hung Wu, Pey-Ching Yang, Von-Wun Soo

Poster Session

The Role of Shared Attention in Human-Computer Conversation 224
Hideki Kozima, Akira Ito

Chinese Word Segmentation and Part-of-Speech Tagging in One Step 229
Tom B.Y. Lai, Maosong Sun, Benjamin K.T'sou, S. Caesar Lan

Corpus-Based Chinese Text Summarization System 237
Jun-Jie Li, Key-Sun Choi

A Study on the Portability of a Grammatical Inference system 242
Hsue-Hueh Shih, Steve Young

Fast Lexical Post-Processing on Cursive Script Recognition 247
Marco A. Torres, Susumu Kuroyanagi, Akira Iwata

The Study of Recurrent Neural Networks for Language Modeling 252
Wen-Jyun Wang, Jyun-Hsiao Lee, Ji-Yi Liu (in Chinese)

A Simple Heuristic Approach for Word Segmentation 257
Wing-Kwong Wong, Chenming Hsu, Jie-Iao Chen, Jien-Chi Yu

Prosody Generation in a Chinese TTS System Based on a Hierarchical Word Prosody Template Tree 262
Chung-Hsien Wu, Jau-Hung Chen

Ambiguity Resolution Using Lexical Association 267
Juntae Yoon, Seonho Kim, Mansuk Song

The Description of the Intra-State Feature Space in Speech Recognition 272
Fang Zheng, Mingxing Xu, Wenhu Wu

Similarity Comparison between Chinese Sentences 277
Lina Zhou, James Liu

Attributive Clauses in Chinese: Theory and Implementation 282
Xiaokang Zhou, Francis Y, Lin

Session 4: Speech Processing

Integrating Long-Distance Language Modeling to the Phoneme-to-Text Conversion Task 287
Tai-Hsuan Ho, Kae-Cherng Yang, Juei-Sung Lin, Lin-Shan Lee

Automatic Speaker Identification Based on Fuzzy Theory and Neural Network Using Genetic Algorithm 300
Ching-Tang Hsieh, Eugene Lai, You-Chuang Wang

*A General Public Application of Pedagogic and Linguistic Vocations of Speech Synthesis: Ordictee 316
Marc Guyomard, Jacques Siroux, Dominique Pernici, Christophe Royer

*A Conversational Agent for Food-ordering Dialog Based on VenusDictate 325
Hsien-Chang Wang, Jhing-Fa Wang, Yi-Nan Liu

*Truncation on Combined Word-Based and Class-Based Language Model Using Kullback-Leibler Distance Criterion 335
Kae-Cherng Yang, Tai-Hsuan Ho, Juei-Sung Lin, Lin-Shan Lee

Session 5: Speech and Language Processing

Recognizing Korean Unknown Proper Nouns by Using Automatically Extracted Lexical Clues 345
Bong-Rae Park, Young-Sook Hwang, Hae-Chang Rim

Logical Operators and Quantifiers in Natural Language 357
Shin-ichiro Kamei, Kazunori Muraki

Chinese Text Compression Using Chinese Language Information Processing 368
Jun Gao, Xixian Chen (in Chinese)

*Combining Multiword Units into a Hidden Markov Model for Part-of-Speech Tagging 380
Jae-Hoon Kim

*A Robust Keyword Spotting System for Mandarin Speech 390
Chung-Hung Chien, Hsiao-Chuan Wang

*A First Study on Mandarin Prosodic State Detection 399
Yuan-Fu Liao, Wern-Jun Wang, Shu-Ling Lee, Sin-Horng Chen

*Rejection in Speech Recognition Based on CDCPMs 412
Mingxing Xu, Fang Zheng, Wenhu Wu

Authors Index 420

Setting Parameters: Triggering and Mis-Triggering

Janet Dean Fodor
CUNY Graduate Center

Abstract

The principles-and-parameters theory of language proposed by Chomsky (1981) greatly simplifies the task of language learning. However, recent research in learnability theory has made it clear that for natural languages there can be no instant "automatic" triggering of parameters. This is because the trigger properties in natural languages are often deep properties, not recognizable without parsing the input sentence.

Current approaches such as Gibson and Wexler (1994) therefore use the sentence parsing routines to identify triggers. Unfortunately, the proposed mechanism for doing so is very inefficient. I show that this is because it does not respect the Parametric Principle: it evaluates millions of particular grammars, rather than establishing the values of 20 or 30 parameters.

By tracing out why this is so, I have found a remedy for it. There is a way of using the parser that does implement the Parametric Principle, and permits efficient learning with no exponential complexity increase. But it calls for a new model of how sentence parsing contributes to learning, and a new conception of parameters and of their triggers: they are one and the same thing, and consist of features or small treelets, made available by UG and adoptable into individual grammars.

This conclusion is in accord with most current theories of syntactic parameterization, including the Minimalist program, HPSG and TAG theory.

References

Chomsky, N., Lectures on Government and Binding, Foris Publications, Dordrecht, Holland, 1981.

Fodor, J. D., "Unambiguous triggers," to appear in Linguistic Inquiry. (Manuscript supplied)

Gibson, E. and Wexler, K., "Triggers," Linguistic Inquiry 25.3, 1994, pp. 407-454.

Encoding Events Across Languages

Beth Levin
Department of Linguistics
Northwestern University

Abstract

All languages must describe the same events in the world, yet the resources that languages have available for expressing certain types of events vary, giving rise to rather different ways of expressing such events across languages. The best known example of this type of cross-linguistic variation involves differences in the expression of motion events --- differences which have long been noted in manuals of translation and have more recently been introduced into the typological literature by Leonard Talmy. This talk will investigate systematic cross-linguistic similarities and divergences in the expression of events, primarily through a case study of the expression of motion events. Taking Talmy's observations as a starting point, I propose that differences among languages can be characterized in terms of the compositional resources available for expressing telic (i.e., bounded) events, rather than in terms of an inherent difference in language type (i.e., Talmy's path vs.~manner languages), as is more commonly done. This approach receives support from cross linguistic variation in the expressions of certain other event types, which can also be understood as involving compositional resources. In addition, I briefly discuss the implications of such systematic divergences across languages for natural language processing, in general, and machine translation, in particular.

Opposing Effects of Word and Character Frequency in the Processing of Chinese Text

Ovid J. L. Tzeng, Daisy L. Hung, Chih Wei Hue

Abstract

Six experiments were conducted to explore the nature of lexical access in the processing of Chinese text. Subjects were asked to perform a lexical decision task under various stimulus presentation conditions. The stimuli were always Chinese words of two characters which could be presented simultaneously or sequentially. Under the sequential conditions, three inter-character-intervals were manipulated to examine the time courses of various frequency effects due either to word or character. Results of these experiments showed a general positive effect of word frequency under all conditions, whereas negative effects of character frequency were consistently observed at ICI longer than 50 msec. These opposing effects of frequency depict a complicated pattern of word recognition which can be explained only by proposing a dual-route model of lexical access in the processing of Chinese text.