読み方の科学：人はどう読み、何を読むのか？
Science of reading behavior: how and what do you read?

相澤彰子　原忠義　Pascual Martinez-Gomez　Chen Chen　冨田恭平　狩野芳伸
Akiko AIZAWA　Tadayoshi HARA　Pascual Martinez-Gomez　Chen Chen　Kyohei TOMITA　Yoshinobu KANO

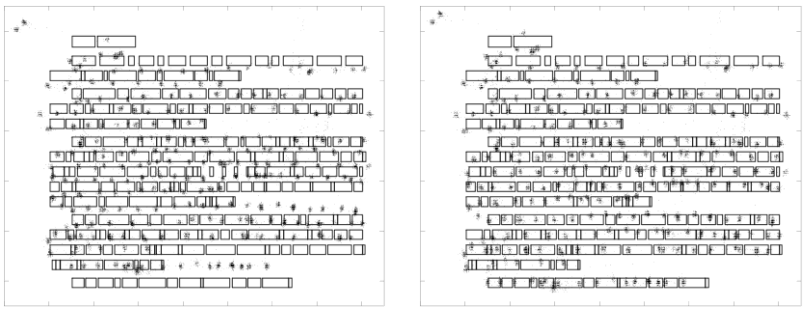
どんな研究？

How does human read texts, and what types of information does he/she obtain to understand the content of the texts? The aim of our research is to answer these questions by matching reader's eye-movement information with semantic structures of sentences, and consequently to achieve deeper text analysis or to bring more useful assistance to human reading. Our presentation introduces ongoing research activities in our group.

Text-Gaze Alignment

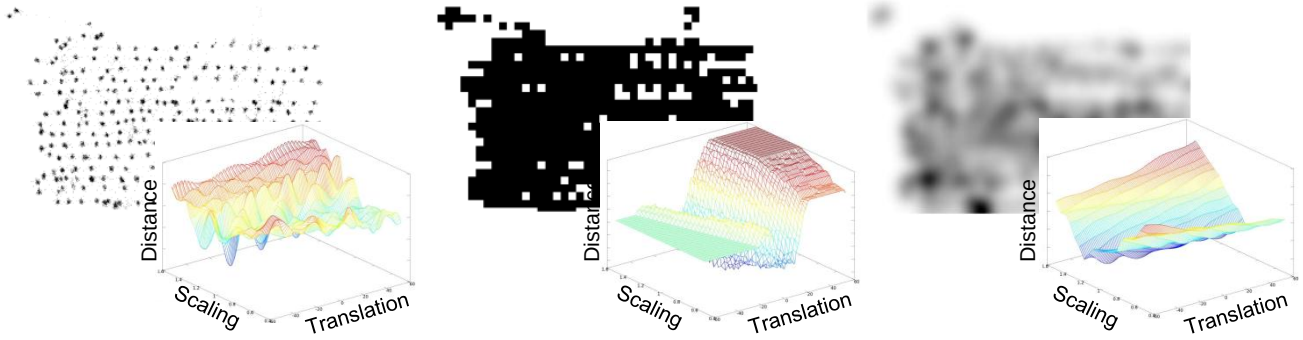
■ Use image registration techniques

Raw gaze data　Linearly transformed gaze data



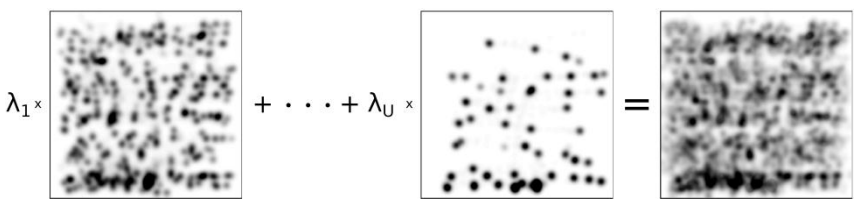
■ Smooth the solution space to obtain the optimal solution

Raw gaze data　Low-resolution gaze data　Blurred gaze data



Synthesize Image Representations of Linguistic Features for Natural Language Analysis

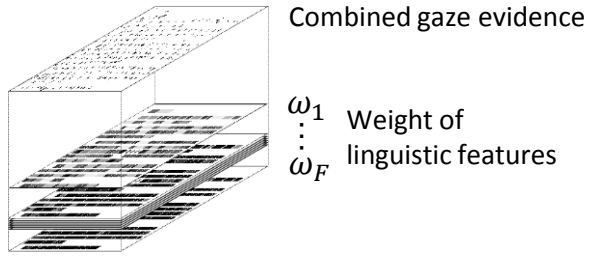
■ Linear combination of gaze evidence from U users with different level of understanding λ_U



■ Image representation of linguistic features

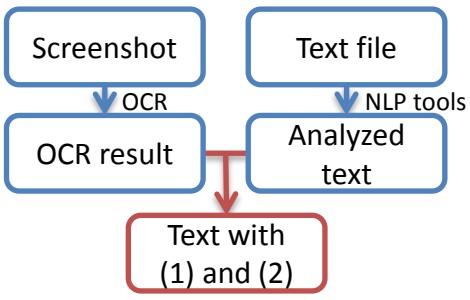
Semantic ambiguity　Nouns

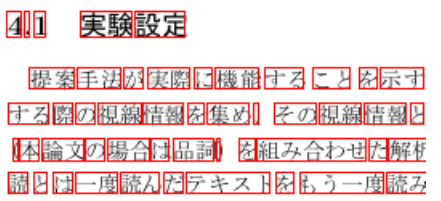
■ Estimation of importance of linguistic features ω_F



Associating Linguistic Features of Words with Display Position

■ In order to analyze gaze behavior in reading text, we should know
(1) Positions of words in the text
(2) Linguistic features of those words
■ Linguistic features: part of speech, dependency, semantic roles of words, …
■ A screenshot and a text file were used to identify (1) and (2)
■ We apply OCR to the screenshot and then correct mistakes of OCR by matching the OCR result and the original text

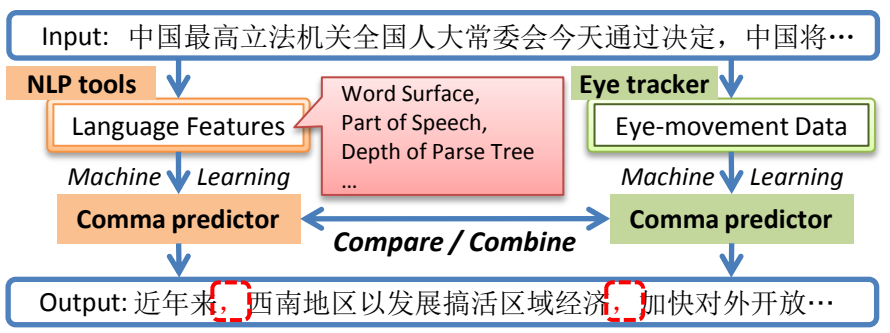




Red boxes are identified words
(Only positions are shown)

Chinese Punctuation Placing for Smooth Reading

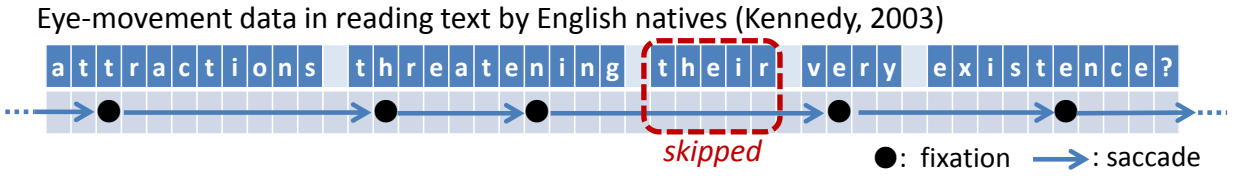
■ Comma in Chinese text plays a more special role than that in other languages
- With/without it, we sometimes have totally different meaning for same characters
→ Contribute more to smooth reading
■ We investigated the comma placement in Chinese sentences by integrating "eye tracking information" and "natural language processing"
→ Create the Comma Distributor/Predictor by using machine learning technology
■ Future work: Find better comma distribution for readability improvement



Prediction of Skipped Words in Reading

■ What in text determines "saccades" in human reading?
■ Predict words skipped by saccades with machine learning
- Clues: word surface, part of speech, length, location, etc.
→ 73% - 84% accuracy for each subject data

Eye-movement data in reading text by English natives (Kennedy, 2003)



NII
連絡先：相澤彰子／ 国立情報学研究所 コンテンツ科学研究系 教授　Email : gazeNLP@nii.ac.jp