A Study of Linguistic Analysis for Classical Chinese Texts

Tomohiko Morioka, Christian Wittern, Koichi Yasuoka
Center for Informatics in East Asian Studies
Institute for Research in Humanities
Kyoto University
Kyoto 606-8265 JAPAN

Naoki Yamazaki
Faculty of Foreign Language Studies
Kansai University
Suita 564-8680 JAPAN

• Classical Chinese texts are different from modern Chinese
  – No spaces or punctuations between words or sentences
  – Different syntax and different usage of words

• Original linguistic analyzer is needed for classical Chinese
  – Basic analysis for classical Chinese
    • Morphological analysis
    • Functional analysis of words
  – Advanced analysis for classical Chinese
    (future work)

• We developed a morphological analyzer
  – Using MeCab, a morphological analyzer engine
    • Original word-class system developed (see right)
    • Dictionary for classical Chinese developed
    • Corpus for classical Chinese under development
  – Please see our demo!

“自立為夜郎侯”

Our Four-Level Word-Class System for Classical Chinese