

# 京都大学自然言語処理ツール

黒橋 禎夫      河原 大輔

京都大学大学院 情報学研究科

## 要旨

これまでに京都大学言語メディア研究室 (旧長尾研究室) で開発してきた形態素解析システム JUMAN, 構文解析システム KNP, 形態素・構文解析結果の修正 GUI, および京大コーパスのデモ・紹介を行う。JUMAN は益岡・田窪文法をもとに, EDR 日本語単語辞書, 独自の連語辞書などを用いて形態素区切りと品詞同定を行うシステムで, 新聞記事ドメインで約 99% の精度である。KNP は日本語文節に関する詳細な文法に基づき, 類似性に基づく並列構造解析等によって文節間の係り受け構造を一意に決定するシステムで, 新聞記事ドメインで約 90% の精度である。京大コーパスは, これらのシステムによって毎日新聞の記事 4 万文を自動解析し, その結果を GUI によって人手で修正したものである。これらのツールはいずれも <http://www-nagao.kuee.kyoto-u.ac.jp/> から入手可能である。

## Kyoto University Natural Language Processing Tools

Sadao Kurohashi      Daisuke Kawahara

Graduate School of Informatics, Kyoto University

### Abstract

Natural language processing tools developed at Language Media Lab, Kyoto University are demonstrated: Japanese morphological analyzer JUMAN, Japanese parser KNP, GUI to modify outputs of JUMAN/KNP, and Kyoto Corpus. JUMAN segments a sentence into words and clarifies PsOS of the words, based on Masuoka-Takubo Japanese grammar, EDR dictionary and a dictionary for fixed expressions. KNP first detects the scopes of coordinations based on parallelism, and then detects the dependency structure of the sentence by a detailed grammar describing functions of Japanese phrases. Kyoto Corpus consists of 40,000 sentences from Mainichi newspaper articles, first analyzed by JUMAN and KNP, then modified by annotators using a mouse-based GUI. All these resources are available at <http://www-nagao.kuee.kyoto-u.ac.jp/>.