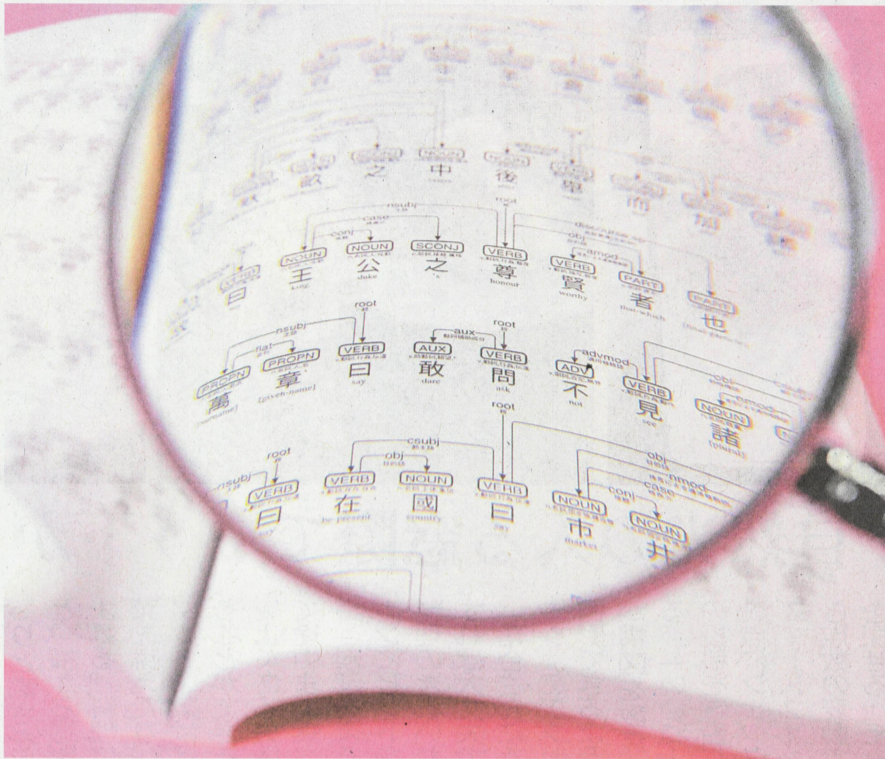


漢文の文法 自動で解析



「孟子」の全文を自動解析し、各単語の品詞や関係性を特定した。500字を超える大部となった

京大人文研 90年の学知

⑦ 安岡孝一

(人文情報学)



やすおかこういち 1965年大阪府生まれ。90年、京都大学大学院工学研究科情報工学専攻を修了し、京都大学大型計算機センター助手に就任。京都大学人文科学研究所付属漢字情報学センター助教などを経て、現在、同研究所付属東アジア人文情報学センター教授、工学博士。著書に「文字符号の歴史―欧米と日本編―(共立出版)など。

からない。このような漢文に対し、あらゆる手法を駆使して、文切りを行い、単語切りを行い、各単語の品詞を解析し、単語と単語の関係を解析し、文における単語の役割を明らかにし、文と文の関係を解析する、というのが研究目標である。

漢文は現在の中国語の文法とは異なり、その構造は未解明の部分が多い。そこで、漢文をコンピュータで解析する上で、どのような文法構造を考えればよいのかが大きな課題である。

東アジア人文情報学研究中心の言語情報学部門では、漢文の文法解析手法を研究している。漢文は、もともと句読点も返り点も打たれておらず、単なる漢字の羅列である。文の切れ目も単語の切れ目も全くわ

単語で区切り 品詞体系グループ化

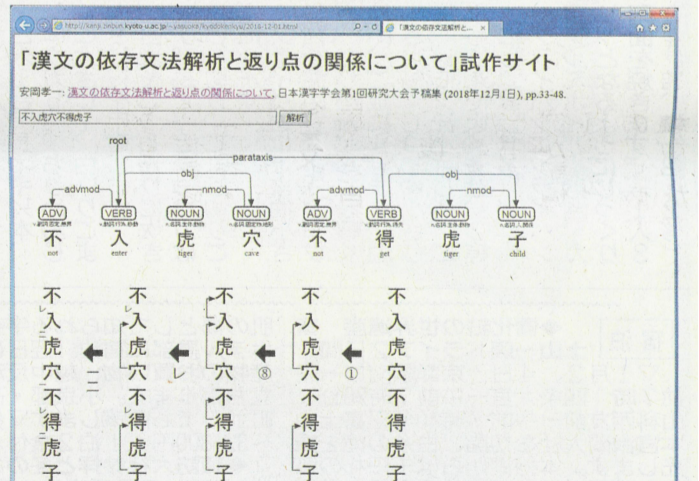
語より日本語に似ている。しかし、動詞と目的語の順序が、漢文と日本語では異なる。英語の文法構造や日本語の文法構造は、漢文には適用できない。

さまざまな言語学的手法を試した結果、1980年代に考案されたメルチュク依存文法が、漢文の文法構造に適用しているのがわかった。この文法は、スラブ語族の比較研究のために考案されたもので、動詞中心主義である。主語も目的語も全て動詞にぶらさがっている、という考え方だ。動詞終構造を記述するのに、かなり都合がいい。

メルチュク依存文法で漢文を記述できれば、次に必要なのは、それをコンピュータで解析するための手法である。プラハのカレル大学には、メルチュク依存文法の改良版を使って、チェコ語などの言語を、コンピュータで文法解析している研究グループがある。このグループが用いている解析手法を、漢文に適用できれば、漢文の文法解析が行えるわけだ。しかし文法構造が似ているといつてもチェコ語と漢文では異なる点が多く、そう簡単には適用できない。

試行錯誤の結果、漢文を単語ごとに区切った上で、各単語に品詞を付与し、さらに、その品詞体系をカレル大学のグループに合わせる、という手法で、何とか漢文の文法解析に漕ぎ着けた。これを表現するために、四書(孟子・論語・大学・中庸)の全文を単語ごとに区切り、メルチュク依存文法の改良版で文法構造を記述し、それらを全て機械学習(いわゆるAI)にかけて、自動で文法解析を行うシステムを構築した。例えば「虎穴に入らずんば虎子を

漢文を入力すれば、自動的に返り点を打ってくれるサイト。<http://kanji.zinbun.kyoto-u.ac.jp/~yasuoka/kyodokenkyu/2018-12-01.html>

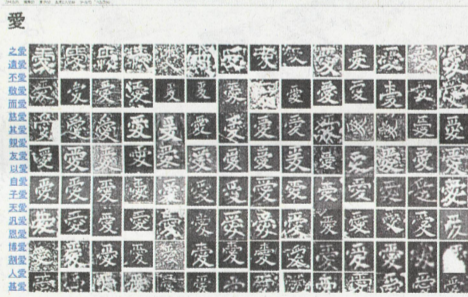


り方など)を解析し、自動的に返り点まで打ってくれるシステムだ。インターネットで公開しているので、漢文を学習中の受験生もぜひ、試してほしい。

得ず」の漢文は「不入虎穴不得虎子」だが、この8文字を入力すると、自動的に各単語の品詞と、その関係性(係

京大人文研の附属施設の一つ「東アジア人文情報学研究センター」は、19世紀以前に漢文で書かれた書籍である「漢籍」の専門図書館であり、資料活用のための多数のデータベースを提供している。

一般にも人気の「拓本文字データベース」は、漢字を入力すると、デジタル化された石碑の拓本にある文字の画像が一覧できる。例えば「愛」という字を打ち



拓本文字データベースに「愛」という字を入力した時の検索結果の一部

漢籍の専門図書館 多数のデータ提供 東アジア人文情報学研

込めば、拓本中にある700以上の「愛」の字がずらりと並ぶ。任意の字をクリックすれば、拓本の原文に当たることもできる。研究者はもちろん、書道の愛好者にも活用されている。

こうした仕組みを研究、構築するためにセンターには四つの研究部門がある。漢籍の本文画像や文字情報のデジタル化を進める文献情報学部門。漢籍独特の分類法「四部分類」に従ったデータベースの構築に取り組む目録情報学部門。龍門

石窟の写真や石碑の拓本など、多彩な文物の画像のデジタル化を担う史料情報学部門。そして、安岡教授率いる言語情報学部門だ。

それぞれが、研究成果を生かした情報発信をするとともに、4部門の連携によって「拓本文字データベース」のような成果が生まれる。安岡教授は「当センターは漢籍の図書館であり、情報を提供することが重要な仕事。研究と公開の一体化が、なよりの特長です」と話す。(阿部秀俊)