人文考 14

京大人文研 創立90年

漢文を読むプログラム　その実力は

安岡孝一教授　（人文情報学）

やすおか・こういち　1965年生まれ。人文情報学。共同研究「東アジア古典文献コーパスの実証研究」班長。

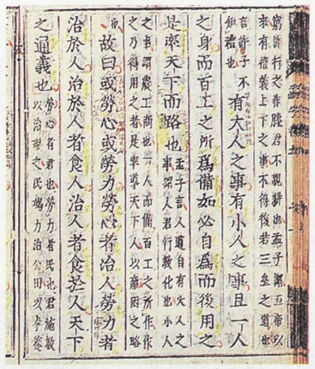
古典中国語（漢文）をコンピューターに読ませるシステムをつくれないか。そんな研究をしていますが、これが非常に難しいのです。

白文（句読点や返り点、送り仮名などのない漢文）には格変化も、現在や過去という時制もありません。それどころか、単語と単語の区切れも文の切れ目もない。漢字がずらずらと並んでいるだけです。

人間が白文を読む場合は、書いてあることの全体的な意味を先に読み取り、うまく通じなければ意味を変えたり、文法構造を考えたりします。しかし、機械に意味といっても困ってしまう。

手がかりになったのは、日本人が訓読に使う返り点です。漢文の基本構造は動賓構造といい、「動詞」「目的語」の順に単語が並んでいます。これが「目的語」「動詞」「助動詞」と並ぶ日本語の構造とは違うので、ひっくり返して読むのです。ということは、返る先と元の間にある動詞を抽出すればいいのでは、と戦略を立てました。

実際の漢文から、どの漢字がどのくらいの確率で動詞や名詞になるかを計算しました。機械学習を積み重ね、いまでは、漢字の並びから単語の区切りや品詞を判別することはほぼできるようになりました。



四書の一つ「孟子」の一節。実際のコンピューター解析では返り点などのない白文のテキストを用意し、品詞や係り受け関係を分析する＝京大人文研東アジア人文情報学研究センター提供

しかし、それだけでは並んだ動詞や名詞などが、どういう風に係り受けするのか文法構造は分かりません。そこで調べた結果、動詞を中心に目的語や主語などの係り受け関係を記述していく方法が有効だと分かりました。

最初は四書の一つ「孟子」を題材に、手書きで文法構造を記述していましたが、途中からは、単語ごとに品詞や係り受けを確率をもとに判断していく手法も用い、機械学習で「論語」や「大学」「中庸」と四書を学習させ、精度を上げていきました。

その結果、漢文の固有の文法ルールを二十数個加えれば、漢文の文法構造はほぼ記述できることがみえてきました。

こうして開発した「UD-Kanbun」という解析プログラムに、センター試験で出題された漢文を白文にして読ませたところ、約6割の精度で文法構造を解析できたのです。現状は、この解析結果に、日本語の語順に並び替える機能と、送り仮名をつける機能を加えるところまで来ています。例えば、「不入虎穴不得虎子」と入力すると、「虎穴に入らずんば虎子を得ず」と出力してくれます。

ただし、課題も浮かんできています。文の中の構造はかなり解析できるようになったのですが、文と文の前後関係をうまく把握することがまだできません。また、機械は固有名詞を知らないため、その部分は読み取れないというのも弱点です。漢籍の専門図書館でもある人文研には、漢文を読むプロたちが大勢いますが、そのレベルに到達するのは、まだかなり難しいのが現状です。

とはいえ、解析プログラムをつくる過程で、漢文に固有の言語構造が見えてきたことで、漢文を比較言語学の対象として捉えることが可能になりました。句点も読点もない融通無碍な表記が19世紀まで使われてきたというのは、やはりそれだけ強固な言語構造があったからだと感じます。

また、「読む」ということを考えると意味の問題は避けて通れません。人間は言葉の外側にある意味も含めて理解していますが、機械にはそれがありません。仮にものすごくうまく意味を説明する機械ができても、受け取る側の人間はそういう機械を気持ち悪いと感じてしまうかもしれないのです。人間と機械の関係性という面でも考えなければならないことは多いのです。

（聞き手・久保智祥）

◆次回は3月25日の予定です。

朝日カルチャーセンター京都教室は、京大人文研の研究者を講師に招く「人文学への誘い」を開いています。3月14日は「日本の近代絵画　西洋とのであい」と題して高階絵里加教授が講演します。午前10時半～正午。受講料は会員2420円（税込み）、一般2970円（同）。問い合わせは京都教室（075・231・9693）。

私のイチオシ本

京大人文研の所報「人文」第66号（2019）に安岡さんの寄稿「四書を学んだコンピュータはセンター試験の漢文を読めるのか」（http://kanji.zinbun.kyoto-u.ac.jp/~yasuoka/publications/zinbun2019-06.pdf）が掲載されている。拓本の文字画像を検索できる拓本文字データベース（http://coe21.zinbun.kyoto-u.ac.jp/djvuchar）の開発にも携わった。

朝日新聞(大阪) 第49642号(2020年2月26日) 夕刊p.4