

インターネットWATCH(16)

OISR-Watch2002年9月11日号

二村 一夫

日本語サイトをハングルで読む——日韓自動翻訳実験

二村 一夫

はじめに

前回と前々回、「ハングルサイトを日本語で読む」と題して韓日自動翻訳サービスを提供しているサイトについてご紹介しました。今回はその逆、つまり日本語サイトをハングルで読むとどうなるか、その実験結果をお知らせしようと思います。発信する側にとっては、こちらの方が気になると思います。

「え?!、君ハングル出来たっけ?」とすぐ突っ込まれるでしょう。自慢ではありませんが、私はハングルはまったく読めません。そんな男が一体全体どうやって日韓自動翻訳の結果を評価するのだろう、と疑問に思われることでしょう。実は私も、最初は誰かハングルの達人にお願いして評価していただくと考えていました。しかし、人を頼まずにテストする簡単な方法を思いついたので、それによる実験です。要するに往復翻訳です。つまり日本語サイトのテキストをいったんハングルに翻訳し、その結果を、再度ハングルから日本語に訳してみたのです。もちろん、翻訳を重ねれば、その結果は悪くなるに決まっています。しかし訳し戻した時に、元の意味を正確に伝えていけば、自動翻訳は実用レベルに達しているとは判断できるわけですから。

I 冒頭文の訳し戻し結果

前置きはこのくらいにして、まずはOISR.ORGのトップページを自動翻訳にかけ、それを訳し戻した結果をご紹介します。利用したのは、All Korea (<http://www.allkorea.co.jp/cgi-bin/allkorea.front>)、それにOCN (<http://www.ocn.ne.jp/translation/>) の二つのサイトで、ともにテキスト翻訳を使いました。

【原文】

「法政大学大原社会問題研究所は、社会・労働問題の研究所であると同時に、利用者の資格を問わない専門図書館・資料館であり、文献情報センターとしても機能しています。1919(大正8)年2月、大原孫三郎によって創立された、社会科学分野では日本でもっとも古い歴史をもつ研究機関です。」

【訳し戻し結果】

☆ All Korea

\$(C!!(B号三二大学オオハなので社会問題研究所は'社会・労働問題の研究所と同時に' ユーザーの資格を問わない専門図書館・資料館であり' 文献情報センターとしても機能しています。

1919(\$C#1#9(B \$(C#1#9(B年年2月' オオハなので長芋であり師父炉によって創立された' 社会科学分野では日本でもっとも古かった歴史を持つ研究機関です。

☆ OCN

法政大学呉オハラ社会問題研究所は、社会・労働問題の研究所であることと同時に、利用者の資格を問わない専門図書館・資料官で、文献情報センターだと言っても機能しています。1919(大正8)年2月、呉オハラマゴサブロによって創立された、社会科学分野では日本で一番古い歴史を持つ研究機関です。

【評価】

すでに前回の韓日自動翻訳ソフトを試した際に分かったことですが、現在の日韓、韓日自動翻訳ソフトの最大の難関は固有名詞、とくに人名の翻訳にあるようです。「法政大学大原社会問題研究所」という名称のうち「大学」と「社会問題研究所」はどちらもきちんと訳しているのですが、「法政」、「大原」で手こずっています。とくにAll Koreaは、大原孫三郎を「オオハなので長芋であり師父炉」と混乱しています。おそらく、これは漢字の一字一字の意味を訳すために起きる問題のようで、日英自動翻訳でも同じ傾向がみられます。もっとも、OCNの場合は日本語の発音をそのまま表記しているようで、訳し戻し結果もかなり原型に近いものになっています。

固有名詞の部分を除けば、冒頭の文章は正確に翻訳されています。「社会問題研究所は、社会・労働問題の研究所であると同時に、利用者の資格を問わない専門図書館・資料館であり、文献情報センターとしても機能しています」という個所をAll Koreaは「社会問題研究所は'社会・労働問題の研究所と同時に' ユーザーの資格を問わない専門図書館・資料館であり'文献情報センターとしても機能しています。」と訳し、OCNは「社会問題研究所は、社会・労働問題の研究所であることと同時に、利用者の資格を問わない専門図書館・資料官で、文献情報センターだと言っても機能しています。」も資料館の館の字を間違えただけで、ほぼ正確な日本語になっています。これなら日本語を知らない方でも、OISR.ORGがどのようなサイトであるかを、理解していただけるのではないかと感じました。

II 見出しの訳し戻し結果

第2レベルへのリンクの見出し、すなわち「お知らせ、ご案内、大原デジタルライブラリー、大原社会問題研究所雑誌、研究所刊行物、OISR-WATCH、社会・労働関係リンク集、労働サイト横断検索、スタッフ個人サイト、関連サイト」を訳し戻した結果はつぎの通りでした。

☆ All Korea

「知らせ、案内、オオハなのでデジタルライブラリー、オオハなので社会問題研究所雑誌、社会・労働関係リンク集、労働サイト横断検索、ステップ個人サイト、関連サイト」

☆ OCN

「お知らせ(通知)、案内、呉オハラデジタル図書館、呉オハラ社会問題研究所雑誌、研究所刊行物、OISR-WATCH、社会・労働関係リンク集、労働サイト横断検索、スタッフ個人サイト、関連サイト」

【評価】

これも大原という固有名詞に関連した言葉を除けば、ほぼ正確に翻訳してくれています。とくにカタカナは、スタッフがステップになったりスタッフになっているだけで正確に訳し戻しています。

III 人名

最後に、自動翻訳でいちばんの難関であるらしい、人名について見ておきましょう。スタッフ個人サイトに並んでいる6人の人名が訳し戻された時、どう変化しているか、興味津々です。

【原文】五十嵐 仁、二村一夫、早川征一郎、鈴木玲、野村一夫、若杉隆志

☆ All Korea

五十嵐一人の夫、後打武に一人の夫、下野カーと三二理で、鈴木ゼロ、野村一人の夫、とキャス気ダツカ市

☆ OCN

五十嵐ヒトシ、ニムだとかズオ、早川ゾングイチで、鈴木領、野村がズオ、ワカスギ タカの時

【評価】All Koreaの名前の間に出てくる「」は全角の空白が化けたもののようです。All Koreaの場合は漢字の意味を翻訳してしまっているのに対し、OCNは人名については日本語の発音を表記しているらしく、訳し戻し結果もかなり原型に近いものになっています。

IV 最後に

このように自動翻訳が実用段階に近づいてくると、われわれ発信者の側でも、自動翻訳ソフトの特徴を知ることが必要になっていると思います。どうすれば、自動翻訳向きの日本語になるのか研究する必要もありますが、まずは、トップページに不要な画像を使わないことが重要でしょう。とくにサイト名をバナーだけで表示したり、テキスト抜きで、ナビゲーション用のリンク・ボタンを使わないといった注意が必要だと思います。固有名詞については名案はありませんが、カッコ内にカタカナかローマ字での表記を加えるというのが、とりあえずの解決策でしょうか。良い智恵がございましたら、お教えください。

(にむら かずお 名誉研究員)

[Last Modified:2000.2.26] `Cannot Open Index File "/home/.sites/28/site/web/cgi-bin/npc.cgi/counter/count.idx"` since 2000.2.15

[OISR-Watch Columns\(Table of Contents\)](#) [次のページへ](#)

[法政大学大原社会問題研究所\(http://oisr.org\)](http://oisr.org)