

・本文書は、「原子力eye」2007年12月号 (Vol.53 No.12) のシリーズ連載「知所先後」の第9回「言語感覚とネット検索」に掲載されたものである。

・同誌編集長のお話によると、シリーズ連載の名称である知所先後(ちしょせんこう)とは、中国の古典「礼記」の中的一篇である「大学」に記述されている「知所先後、則近道矣」から採ってきたものとのことである。意味は「先にすべきことと後にすべきことを知ることが(知所先後)、近道なのである。(則近道矣)」と言う。

・同誌の原文の記事について一読されることをお勧めします。

【問合先】(株)日刊工業出版プロダクション 原子力eye 編集部

〒103-8548 東京都中央区日本橋小網町14 - 1

Tel:03 - 5641 - 8355/Fax:03 - 5641 - 8360/URL <http://pub.nikkan.co.jp/mgz/eye/index.html>

「言語感覚とネット検索」

NTTアドバンステクノロジー(株) 印牧直文

1. ドリームキャッチャー(dreamcatcher)との出会い

米シリコンバレー駐在時代に暗号理論の専門書を読んでいたところ、その中にナバホ(Navajo)のインディアンの話が出ていて驚いた。その驚きがきっかけとなってナバホ族の居留地(Indian reservation)であるモニュメント・バレー(Monument Valley)に家族と旅をしたことがある。ナバホの居留地はナバホ・ネイション(Navajo Nation)と言われ、アリゾナ州、ユタ州、ニューメキシコ州の三つの州にまたがる広大な地域で、強い自治権を持っている。その中心地の一つがモニュメント・バレーである。岩山が砂漠の地面からよきによきと空に向かって突き出ているような雄大な光景である。テーブル型の台地がメサ(Mesa)、石柱のような岩山がビュート(Butte)と呼ばれている。これらの岩山を見ながら、ナバホ族は砂漠地に建てたドーム型家屋のホーガン(Navajo hogan)に住んでいる。確か、東側の窓近くだったと思うが、ドリームキャッチャー(dreamcatcher)という幸運を呼ぶ装飾品を吊るす。

2. インフォメーションキャッチャーと言語感覚

ドリームキャッチャーは木の輪の中に蜘蛛の巣状の目の粗い網(Web)を組み込んだ装飾品である。夜中に侵入する悪い夢をこの網で採り、良い夢は網を通り就寝中の頭に入り込むと彼らは信じている。この装飾品のオリジナルはオジブワ族(Ojibwa)のものとされ、その解釈は異なる。ベッドの上に吊るし、良い夢をこの網に引っ掛けて採り、良い夢が就寝中の子供の頭に入り込み、悪い夢から守るという解釈となっている。これがドリームキャッチャーの名前の由来である。インターネットを使って頻繁に情報を検索している筆者が、Webを使って適切な情報だけをキャッチするインフォメーションキャッチャー(情報検索手法)を作りたいと考え始めたのはこの頃である。情報の世界では「インディアンの夢」は「概念」に対応し、このため「概念」を表現する単語や文をどのようにキャッチすればよいかという着眼点に転化される。単語や文は、言語の種類によって「概念(定義)」が微妙に変化する。言語感覚は情報検索手法では重要な要素の一つとなっている。

3. 暗号通信兵、ナバホ・コードトーカー

モニュメント・バレー題材の映画として、西部劇の駅馬車、バック・トゥ・ザ・フューチャー3、イージーライダー、ウインドトーカーズ等が知られている。ウインドトーカーズ(Windtalkers)とは、第二次世界大戦のサイパン島や硫黄島の戦いで実際に活躍したナバホ族の暗号通信兵であるコードトーカー(code talker)の愛称である。1968年まで米軍極秘事項として伏せられていたが、その後公開され、1982年にロナルド・レーガン大統領に表彰され、2000年にナバホ・コードトーカーに議会名誉黄金勲章が授与されている。コードトーカーは、最前線で戦況を把握し、援軍に伝達する英単語をそれと同じ文字で始まる別の英単語に置き換え、更にそれをナバホ語で翻訳して援軍に情報伝達する通信兵であり、コンピュータ暗号処理の基礎的な考え方となっている。

4. 文化を加味するローカライゼーション情報検索

暗号とは情報伝達相手を限定するローカライゼーションの最たる言語の一つである。代表的なローカライゼーションは、使用言語の違いによる地域性である。この地域性は、地理学上の物理的な地域性(言語別話者分布)とインターネット上の論理的(バーチャル)な地域性(ネット利用者分布)に分けられる。更に、同一言語においてもローカライゼーションが多数存在する。それは文化を持つ集団である。文化は独特のコミュニティや群れを作り集団帰属の有無を確かめ易くしている。まず社会文化が挙げられる。例えば、「ナウイ」という単語を使う社会文化が20年以上前にあった。この頃、単語「ナウイ」を使った情報が新聞・雑誌等のデータベース(DB)に現在蓄積されているが、この単語を知らない年齢層はこの単語が頭に浮かばず検索の非効率化が生じてしまう。「ナウイ」は新鮮さや格好よさを意味するが、その後の社会文化では「イケテイル」という単語に変化し、「ナウイ」を知っている年齢層は「イケテイル」という情報検索が難しくなる。このような新語や流行語を考慮するとともに、外来語や造語のカタカナ語も考慮するローカライゼーション情報検索が重要になる。

5. 2つの言語を使ったダブルチェック・ネット検索

ビジネス文化もローカライゼーションの一つである。例えば単語「インターフェース」は経済産業省とその管轄下の外郭団体や施設では多く使われている。ところが、旧郵政省(現総務省)とその外郭団体や施設では「インタフェイス」という単語を使う文化となっている。後者は伝統的に長音記号「」を嫌う文化と言われている。ビジネスでは契約書(購入仕様書)の提出とこれに対する契約金額の受け取りが一对になった活動が複雑に絡み合っている。これらの書類で使われる約束事の「概念(定義)」を誤判断すると多額の損失を被る。ここにネット検索の必要性が発生する。社会文化やビジネス文化のローカライゼーションを加味したネット検索による契約書類の確認作業が必要になるのである。異なるビジネス文化を通して微妙な概念と論理思考を知るためには、同一情報源に対して異なる記者(概念)が異なる言語(思考)で記述した2つの記事情報を併用する分析方法が有効である。この種の簡易手法としては、ネット上のフリー百科事典の英語版と日本語版を併用する方法がある。