

国立国語研究所学術情報リポジトリ

世界の言語研究所 (3) インド国立科学ドキュメン
テーションセンター INSDOC (インド)

メタデータ	言語: jpn 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/2009

インド国立科学ドキュメンテーションセンター INSDOC (インド)

Ashok K. Chawla (INSDOC)

1. はじめに

INSDOC は1951年にインド政府が働きかけて国連の資本・技術援助のもとで設置されたインド国内のドキュメンテーション活動の中核機関である。1952年にインド政府が当機関を Council of Scientific & Industrial Research (CSIR) 傘下の40ある研究機関の一つとし、最初の段階では、同 CSIR の国立物理学研究所 (National Physical Laboratory) の中に設置された。そのときも情報処理とドキュメンテーションが大変重要な分野として認められていたので、その後すぐ独立的な情報提供の先駆機関となった。スタートの時点ではドキュメンテーション、翻訳、出版、複写、印刷そして入力 (タイピング) の仕事を中心にしていたが、現在は情報科学分野の国際組織として科学技術に関するライブラリーデポジットと自動ライブラリー情報サービスのインテリジェントシステム開発を重点的に行っている。定員は現在285名であり、そのうち研究員は108名である。研究の年間予算はインドルピーで約 45 million (現時点で1ドル39ルピー) である。

情報処理になるとまず第一に問題となるのはことばであった。元々、インドにおける言語研究は非常に複雑な仕組みとなっていた。その背景にあるのは膨大な人口と数千年の歴史の中での外国語の影響である。古いことばベータ語で書かれたインド最古の文献「リグ・ベータ」と科学の分野に関する文献はインドのあらゆる方面の学者や外国人研究者の間で大変興味深い研究の対象となっていた。昔の文献がまだ完全にサンスクリット語に訳されていないうちに、サンスクリット語は南インドのドラビディアン語と混ざり、その結果いくつかのことばが生まれた。インド政府外務省のある PR 資料によると現在インドでは844種類の違う言語が使われているようで、そのなかでメジャー言語は15種類となっている。中央政府の二つの公用語ヒンディ語と英語の中でヒンディ語はインドの約45%の人が母国語又は第二言語として使っているのだが、国語とはなっていない。州によってことばが違う場合があり、ヒンディ語があまり使われていない州もある。また、ヒンディ語を第二言語として使っていないところは英語が第二言語となっていて、英語がヒンディ語の次に国の共通なことばと言われている。

1950年代の前半にサンスクリット語又はその専門用語だけを国全体で共通に使われる可能性を探るプロジェクトがスタートしたが、その結果もでないうちに科学・技術関係の文献も教育もほとんど国中に英語で行なわれるようになり、INSDOC で使われる言語も実質上は英語だけとなった。したがって、INSDOC で使われることばは英語となったが、翻訳部に限っては外国語から英語又は英語から外国語への翻訳に加えて一部インドの他のことばも対象となった。

2. INSDOCの使命と情報サービス活動の位置づけ

INSDOC の使命は次の通りである。

- ・ 様々な科学技術分野の情報の収集とその開拓
- ・ 科学技術分野の団体への情報提供
- ・ 国内外の情報システムとそのサービスにおいて適切な関係を結ぶこと
- ・ 国内で発生した全科学技術情報の保管
- ・ 科学技術情報の研究開発を行うこと
- ・ 科学技術情報サービスのための人材育成
- ・ 国内に増加する情報ネットワーク資産に必要な技術管理の適切な運用と採択

これらの使命を果たすために INSDOC が科学技術情報保管機能として CSIR 傘下に設置され、インド国内の科学技術コアジャーナル（主な雑誌）と重要な科学技術の単行本を NSL（国立科学図書館：INSDOC 本体内）で収集し、全インドの R&D 機関、研究所等をサポートできるようにいくつかの情報サービスを提供する。情報サービス活動は主に情報の収集、分析、翻訳、印刷、提供に分けられる。外国語で書かれたものの翻訳は昔から大変重要な業務となっていたのでそのための研究員の数も総研究員の約半分だったが、この10年の間に翻訳の仕事は正式の研究員がやるよりも外注ですます傾向が強くなりつつあるので翻訳業務だけにかかわる研究員の数が少なくなっている。内部の人は外の人ややったもののチェックやその他の管理の他に外部からの研修生にソース言語の解釈とターゲット言語の表現の仕方の訓練をやる義務を負う。従って、研究員は翻訳用のサポートシステム作り、つまり教材・素材の準備、あらゆるツールの購入・作成、訳すテクニックの研究に力を入れながら、言語の文法的な面、辞書作成、外国語と英語のペアの対照研究、機械翻訳に関する研究も行なっている。

3. 情報活動のなかでの言語に関する流れ

INSDOC の1955年度の年報によると当時は主に10種類の外国語の英訳の仕事が主流で、その量をみると一番多いのはドイツ語でその次ロシア語、フランス語、日本語、イタリア語、スペイン語、オランダ語、ポーランド語、スウェーデン語そしてハンガリー語の順で注文に応じて仕事を行なうことが主流だった。ユーザーは CSIR のすべての研究所と政府の各省庁また、国営の防衛研究所などだった。1980年代の後半からの傾向をみると、ロシア語科学情報センターが設けられ、アメリカからもロシア語関係の特別カバーツーカバー翻訳プロジェクトがあって、ロシア語の翻訳の仕事が増えた。民間会社と違って翻訳が学問活動として行なわれるので、結果としてロシア語関係の研究が一番さかんになった。したがって、ロシア語専門家が増えた。しかし、最近ではロシア語の活動が減ってきている。そして、ドイツ語に並んで日本語がトップにあがっている。その背景には、一般にドイツと日本から多くの会社がインドに進出していることに加え CSIR の全ての研究所でもドイツと日本の雑誌、論文が頻繁に引用されているからである。その需要に答えるためにも1970年代に行なわれた機械翻訳プロジェクトを復元させる動きもあったが、現在市販されている機械翻訳システムを使おうとしてもまだまだ翻訳に役立つものがないため、重点は翻

訳サポートシステムの作成の方に移っている。

もちろん、近年各情報機関ではコンピュータの進歩や研究開発の高度化により、機械翻訳が実用化されてきている。しかし、これらは翻訳文の入力から出力まで機械によって完全な翻訳文を得られるものではなく、人間が段階的にプレエディット、ポストエディットすることで生成させる方法が進められているのが現状である。

翻訳をしながら研究成果をだすことが期待されるので機械翻訳よりも実際的な問題に関わるテーマに人気がある。例えば、日本語の場合は固有名詞の読み方、古い文献と最近の文献における専門用語、一般用語、外来語、文体と助詞などの使い分け、論文タイトルの日英対照研究そして学問的な論文と工場の現場の人が書いた報告書などでのことば使いの違いなど。仕事の分担は、現在所内にいるドイツ語、日本語、フランス語、ロシア語、ポーランド語、スペイン語、イタリア語そして中国語の専門家がそれぞれのことばに関する仕事と研究を管理していて、その他のことばは基本的には外注となる。日本語を含めていくつかのことばの通訳サービスもあるが内部の人のできる仕事しかひきうけないことになっている。仕事と研究の内容から機械翻訳のコーパス作り、外国語で出版されている1次ジャーナルからの専門用語の収集などは日常の仕事として行なわれている。

4. 日本語に関する研究

今まで日本語に関しては化学の日英ヒンディ対照研究、特許用語の研究、複合語における要素と要素の関係についていくつかの論文がでていますが、最近、特に技術者向けの日本語教育と翻訳サポートシステムの開発に力を入れている。日本語に関する研究員が2名で約15年間で行なった研究を簡単に次のように紹介できる。

1980年代の前半では主に専門用語の諸問題と日本語をスムーズに分析できるように文脈と文法についての研究が主流だった。その結果として化学と鉄鉱関係の専門用語の収集、分析をし日英と日ヒンディ対照研究を行なった。次に、内容がコンピュータの専門用語に変わり同じ対照研究が行なわれた。成果として対照研究の論文がいくつかあるが辞書をつくるまではいたらなかった。1980年の後半では人手による翻訳が機械翻訳にとってかわられるという深刻な問題がでてきたときに国連とインドの科学技術省がリードをとって機械翻訳に関する国際会議を開き、ドイツの民間会社 Seimens も入れて (Seimens の機械翻訳システムはすでに商品化されていて、ミスは10%ないし30%程度) アジアのいくつかの国の専門家が参加した。

INSDOC では機械翻訳に関しての調査にはすでに入っていたが、主にはドイツ語と日本語のいくつかの機械翻訳ソフトを使って、単純な機械翻訳と人間が手を加えた機械翻訳、初心者からの翻訳それから当時の INSDOC の最終的な翻訳を比較したものを発表し、国際会議の結果として Society for Machine Aided Translation (機械を使った翻訳に関する学会) が生まれた。

これらと平行して日本語の電気工学の複合語における要素と要素に関する研究をテーマにしていくつかの実験を行なった結果、INSDOC では Translation Support System の開発に力を入れるようになった。そして、機械翻訳と少し違う概念でこのプロジェクトは普通の仕事をしながら

ら完成する方向で動いている。基本的な内容は文，語，要素の三つのレベルでのデータ収集，分析をやる一方で文脈の理解のためにパターンのデータベースを作って全ての情報をなるべくオンラインで使えるようにする考え方である。

また，この数年インドから日本への人材派遣が増えているので日本語教育用教材作成はもう一つの大きな活動になりつつある。そして，論文タイトル，表現，用語そして要素レベルの日英対照データベースの作成の他に公用語のヒンディ語との対照研究がこれからの課題である。日ヒンディ対照研究が本格的に始まるとインドでの日本語研究も大きく影響されると思われる。

Indian National Scientific Documentation Centre (INSDOC)
14, Satsang Vihar Marg, New Delhi-110067, India
Homepage: <http://www.imsc.ernet.in/~insdoc.html#em>
Email: chawla@sirnetd.ernet.in