

談話室

外国語翻訳データ・バンクの構想

披田野 陽一*

科学技術英文作成のための語学情報を収録した、外国語翻訳データ・バンクの設立を提案する。

従来は自己の所有する科学技術を海外諸国に売込むのに消極的であった企業にも、オイルショック以降は様変わりに、経済的交流に限らず、科学技術的交流をも積極的に押し進めようとする姿勢が生まれている。特に英語圏への接近は著しく、今後、ますますその勢を増すことは間違いないように思われる。

このような状況のもとにあって、科学技術者にも交流の一端の担い手として、十分な英語力が要求されているが、環境に恵まれ特殊な訓練を受けた少数の者の英語以外は、残念ながらほとんど使いものにならないのが実情である。

しかし、この緊急の必要性から、科学技術者も科学技術のみならず英語についても、完全な「遣い手」とまではゆかなくとも、最低書く力位は身につける必要がある。そこで、自分の英語レベルと実用レベルとの差をどうすれば埋めることができるか、を一度は真剣に考えてみる価値はあると思う。これからは、この答によって科学技術のPRや交流の担い手として実力を有するか否かが決まる、とさえ言えよう。

翻訳作業は人間の営む創作作業の一面であり、同時に記憶を最も有効に活用する作業でもある。もし記憶の補助手段を用いて語学情報を蓄え、必要な時に必要な情報を入手できるならば、王道とは言えないまでも上記問題の解決に役立つと考える。翻訳の専門家であれば、この情報の準備は常に怠りなく行っているが、即席の「両刀遣い」にこれを期待することは無理である。そこで、自然言語や検索法など情報科学の最新成果、さらに実際の翻訳作業プロセスを分析した結果を十分に取り入れた、合目的な内容を有する外国語翻訳データ・バンクの設立を提案する。以下、その概要を翻訳データの内容、科学技術部門、完成に要する期間と資金の点から、簡単にのべることにする。

翻訳データの内容：文章表現作業は、表現したい概念の主体を明確にし、その概念に関連する言語記号の名詞群、名詞相当語群の中から、主体を表現するのに最も適した語を選択し対応させることから始まる。重要な点は表現という文章定着作業が、名詞の選択、対応付けで始まること、およびそのとき名詞が最も重要な結晶核の役割を果たしていることである。そこでまず、科学技術上の概念の主体となる名詞の多面的な表現の様相をデータ化する必要がある。その様相の具体例を以下にあげる。

- (1) 概念の主体としての科学技術用語（検索語兼用）
- (2) 相当日本語とそのローマ字綴り（外国人向）
- (3) 科学技術用語概念の説明（科学技術英文作成は技術専門家のみでなく、弁理士の米国出願特許作成にみられるような場合も多いので、そのためにこれがあると役立つ）
- (4) 科学技術用語群間の概念上の上位、下位の階層性
- (5) 同じく用語群間の同類異類対立関係などの関連性
- (6) 概念と造語法からみた複合語の様態や構成法（複合語の可能性の示唆）
- (7) JIS や学会による制定用語の指示等。

すなわち、科学技術用語群を主要な検索語を中心に集大成したものである。さらに言語世界は民主的世界であるから、概念の主体を特定用語で統一できない面があり、類語が増加する。そのために類語対策として索引も必要である。その国の科学技術力は、その用語が如何程多く、その国の国語で表現できるかによって推測できるのではあるまいか。それ故に、科学技術庁が科学技術用語シソーラス日本語版の製作を行っている（日刊工業；S47-10-16）ことはまことに喜ばしい。

つきは自分が選択した用語の慣用性を示し、かつ文章作成の際の英借文の見本となる用例が必要である。

* (株)日立製作所横浜研究所

経験から述べれば、それは「・」区切りの長さを最小単位とすべきである。通常の語法辞典は英語に堪能な者には記憶の確認のために役立つ、初歩の者には語義調査のために役立つが、非常に人数の多い中級レベルの者にとっては不十分である。その理由はこれらの多くの辞典が最も簡単な単文表示、あるいは to have 式の表示法を採用しており、幅の広い前後関係等、真に必要な語学情報が十分に示されるとは思われないからである。日常事についての英文作成でさえ、英借文と言われている。まして、専門分野が異れば同一用語でも関連動詞、形容詞の異なる例が多い科学技術文の表示法には十分な工夫が必要である。そのために、自分の選択語を文章全体から判断する基準用例、また論理展開で重要な接続・修飾表現や語彙結合の慣用性、会話語と文章語等の混合のない文体上の同質性を示す用例は、「・」単位の表示法で始めて可能であると考えられる。

科学技術部門：一言で科学技術と称しても、これは基礎部門の理学から応用部門の工学、さらに境界領域まで含まれるので、いわば人間存在に係る広範な部門からなる巨大な王国である。これを学問的体系にしたがって厳密に分割するのではなく、科学技術英語のもつ実用性を重視して分割するならば、科学技術についての 12 部門と全体に共通な事項をまとめた 1 部門、計 13 部門に分けて考えることができる。すなわち、

第 1 部門＝理学・応用理学(25,000)、第 2 部門＝鉱物・材料科学(8,100)、第 3 部門＝機械(17,000)、第 4 部門＝電気(8,800)、第 5 部門＝工業化学・医薬品(微生物工学含む；10,400)*、第 6 部門＝農林・水産・畜産(生物工学含む；4,100)*、第 7 部門＝土木・建築(2,000)*、第 8 部門＝環境・保健科学(8,000)*、第 9 部門＝システム・情報科学(5,400)*、第 10 部門＝工業所有権(1,100)*、第 11 部門＝海外技術交流・広報技術(商業英語含む；2,200)*、第 12 部門＝管理・行動科学(時事英語含む；5,800)*、第 13 部門＝共通(4,400)*。

この部門分割は TEST(米国工学会議編)、NASA シソーラス(NASA 編)、技術士試験要綱(通産省編)科学技術翻訳士会要綱(科学技術翻訳士会編)、UDC(英国規格協会、日本ドキュメンテーション協会編等)を参考にした。各部門はさらに小さな科学技術分野に細分される。ちなみに、電気部門は法規類まで含めて 26 分野である。また、() 内の数字は TEST、NASA の分野を上記部門に再編成して求めた部門毎の基本用語の概数である。ここで用語は重複して数えたが、そ

れは用語の中に、部門間で借用されて他部門に移り、そこに帰化したものがあるからである。また、上記二資料は特定部門を特に充実した様子が感じられるので、それ以外の部門にはその部門固有の用語がなお相当数有ることが推測される。そのことを*印で示した。この項の最後として、部門毎の翻訳データの規模は、説明は省略するが各部門の用語数に比例した量と考えてよいと思う。

完成に要する期間と資金：翻訳データは検索の便を考え、カード(B6)方式を採用する。追加語を含めて、検索語カード数は用語総数の 1.2 倍と見ると約 12 万枚、用例カード数を加えたカード数は約 120 万枚となる。実行可能な作業能率(実証済)は検索語数にして 400 語/人・年(用例カード製作含む)であるので、15 名編成で作業するとして、完成するのに 20 年間掛かることになる。また資金額はカード印刷費、資料購入費、作業費等を含め、現貨幣価値で約 6,000 万円、年平均約 300 万円が必要である。この中には不明の経費は含まれない。なお、上記総額の約 10~15% はデータ作成開始時に計上することが必要である。ところでこの金額は利益を産まない資金としては高額であるので、一企業のみで負担することは難しいのではあるまいか、そのために一部門完成するたびに、その部門の辞典類や教科書、参考書を製作してその利益をバンク設立費用として補充する事はどうだろう。これは出版界の実情を知らない甘い考えであろうか。科学技術英語を学習するための優れた資料も、翻訳データ・バンクとは別に非常に必要である。幾度となく議論されているように、日本の英語学習は時事英語や商業英語等のそれを除いて、概して教養主義的であり、旧来の工業英語はこの流れの延長線上にあると思われる。これでは不十分であるので、科学技術に造詣の深い者が本格的にこの分野の英語研究に携わる必要性を痛感する。

個人的関心から、休日を利用してこの問題に着手し、すでに 6 年になる。外国では国家事業として、この種の機関を設立しているが、その特殊な性格から、日本でも公的機関として設立する必要がある。英文と日本語の機械翻訳技術(?)を一企業人の感覚で研究していて、その前準備として生の言語データを蓄積した本データ・バンクの必要性を感じた。この翻訳データ・バンクが、近づく 21 世紀の人達へのささやかな贈物とならんことを願っている。

(昭和 50 年 4 月 22 日受付)

(昭和 50 年 6 月 27 日再受付)