

機械翻訳システムの専門用語について

田中 康仁
兵庫 大学

E-mail: yasuhito@humans-ke.hyogo-dai.ac.jp

〔 0 〕はじめに

機械翻訳システムが多く企業で開発されている。だが、このシステムは日本語の辞書、相手国語の辞書、対訳辞書等さまざまなものを準備しなければならない。そのためなかなか品質が向上しない状況である。そこで品質向上を計るために専門辞書に焦点をあてて考えてみる。

〔 1 〕機械翻訳システムの専門用語

ある企業の機械翻訳のカタログを調べてみると、次のような専門用語があると宣伝している。

専門分野	登録語数
	170,000
情 報 処 理	17,800
電 気 / 電 子 / 通 信	16,600
土 木 / 建 築	13,500
自動車/鉄道/船舶/航空	28,000
自 然 科 学	24,600
生 物	8,800
機 械	21,700
工 業 化 学	36,700
ビ ジ ネ ス	60,000
合 計	230,000

(A社カタログより)

しかし、この程度の専門用語で充分なのだろうか？ このカタログには分野別専門用語が使用されている専門用語の何%をカバーしているという説明はない。

また別の機械翻訳システムを作成している企業のカタログを見ると24分野で120万語の専門用語を用意している。専門用語の語数が多いことは我々に安心をもたらす。だが、これにもどの程度を満たしているとは書いてない。専門用語を利用した翻訳例がのっている。

ビジネス用語辞書を使用

英日
翻訳

【原文】 This company has annual sales of 2 billion, and its earnings per share is high.

【基本】 同社には、20億の例年の販売があって、そのシェアあたりの収益は高い。

【ビジネス】 同社には、20億の年間売上高があって、その1株当たりの利益は高い。

日英
翻訳

【原文】 雇用保険法が改正され、給与の約25%が休業給付として支給されることになりました。

【基本】 The employment insurance method was revised, and about 25% of the salary was provided as a closed supply.

【ビジネス】 The Employment Insurance Law was revised, and about 25% of the salary was provided as a disability benefit.

† Machine Translation System
- Technical Term -
Yasuhito Tanaka
Hyogo University

この会社の専門用語の辞書を用いれば、このように全ての文にうまくあてはまり、正しい訳文が作り出されるのだろうか？ 逆にこれはたまたま、うまく訳せた例文ではないだろうか。この点も調査しなければならない。どの程度有効であるかわからない。

〔 2 〕専門用語の交換性について

A社の機械翻訳システムからB社の機械翻訳システムに変えようとする場合、これまでに開発した用語辞書がうまく移行ができるような方法が考えられようとしている。

アジア太平洋機械翻訳協会に加盟する機械翻訳メーカーが中心となり、各社のMTシステムのユーザー辞書と相互変換可能なユーザー辞書共通のフォーマット (Universal Plat Form;UPF) の設計を開始し、仕様の公開を行い始めている。

しかし、専門用語を供給している企業は出版業界である。出版業界は紙の本からCD-ROMの本へ移行することは考えているし、そのようなものが出現している。

このCD-ROMがうまく機械翻訳システムにすぐ利用できるようにはなっていない。もっと出版業界の知識を利用する方法も考えなければならない。

この状況はコンピューターが出現してまもない頃、あるメーカーのコンピューターを導入すると全てのものがその会社の装置でなければならなかった。しかし、I/Oチャンネルの標準化や技術公開により、今では他社の周辺機器を自由に接続することが可能になった。

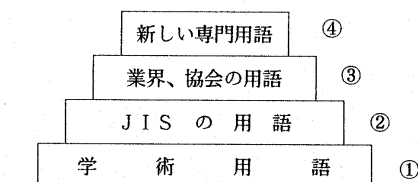
出版業界の専門用語辞書もこの機械翻訳システムとも自由に結合できるようにしてほしいものである。このようにして専門家の知識がうまく利用できる方法を考えるべきである。

専門用語辞書接続の技術公開と各社の統一規格の作成が重要である。

〔 3 〕専門用語の階層性について

専門用語は多くの専門家がある特定の分野の概念を表現するために作られたものである。

それゆえ、定義が定まった基礎的なものから、基礎的な専門用語を組み合わせた複合的な専門用語まで幅広く存在する。それらは次のような階層をなしている。



① 学術用語は文部省を中心に各専門分野の人々、学会等が中心となり制定されたものである。2,2の分野が

あり各分野には次のような用語がまとめられている。

- | | |
|---------|------------|
| 1) 数学 | 12) 建築学 |
| 2) 天文学 | 13) 船舶学 |
| 3) 物理学 | 14) 航空工学 |
| 4) 気象学 | 15) 計測工学 |
| 5) 地震学 | 16) 原子力工学 |
| 6) 分光学 | 17) 歯学 |
| 7) 化学 | 18) 論理学 |
| 8) 動物学 | 19) キリスト教学 |
| 9) 植物学 | 20) 図書館学 |
| 10) 遺伝学 | 21) 土木工学 |
| 11) 機械学 | 22) 採鉱冶金学 |

これらの学術用語を基礎として、日本工業規格 (J I S) が制定されている。J I S は科学技術分野が中心である。

J I S の専門用語分野は次の通りである。
土木・建築／機械／電気／自動車／鉄道／船舶
鉄鋼／非鉄金属／化学／繊維／鉱山／パルプ・紙
窯業／医療安全用具／航空／情報技術／その他
〈包装、溶接、原子力、放射線 (能)、品質管理、音響、照明、他〉

- 各分野で制定された用語規格及び重要個別規格、284規格に規定された最新のJ I S用語約73,000語を収録。
- 各用語には読み方、対応外国語 (英・独・仏語)、定義、慣用語、略号、記号、量記号、単位記号及び図・表を収録。

(日本規格協会カタログより)

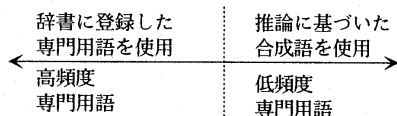
最新のJ I S工業用語大辞典には7万3千語収めている。

さらにこれらを基にして業界、協会、団体が専門用語辞書をまとめている。専門用語辞書にまとめる前の段階では、業界の通用語のようなものが確立しつつある。

〔4〕確立した知識と推定により合成された知識

専門用語の収集、整理は大変な時間、労力、費用がかかるものである。収集した専門用語の適合率も調べなければならないが、調べられたとしても適合率はなかなか向上しない。そこで基礎的専門用語を用いて、ある規則を使い合成して派生的な専門用語の訳語を作り出そうとしている。私の考えであるが専門用語はなるべく大量に集め、それらで対応できないものは推定により合成された専門用語を使うべきである。推定により合成された専門用語の訳語にはどうしても誤りが含まれる。

それゆえ、専門用語の使用頻度のうち80%~90%は収集整理され、確立した専門用語を使い、残り10~20%は推定と合成による専門用語を使うべきだと思う。そうすれば推定と合成専門用語の訳が80%だとしても、全体として専門用語の訳の誤りは数%程度となる。



〔5〕専門用語の適合度

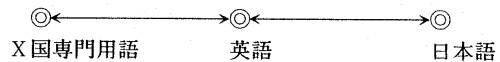
専門用語の翻訳システム上での適合度を調べるにはどのようにしたらよいか考えてみる。我々には色々な制約が多くあるが、それらのものには束縛されないとした。

- 1) 大量の専門分野の文書データの選択と入力
又は機械可読ファイルを見つける。(パラレル・コーパスが望ましい。)
- 2) 1) のコーパスから専門用語の抽出
種類と頻度を付ける。
- 3) 機械翻訳システムに翻訳させる。
- 4) どの程度正しく翻訳されているか否かの判定を種類別、頻度別に行なう。
- 5) 正しく翻訳されていないデータを集め、どのような専門用語を追加すればよいか調べる。

このような1)~5)までの工程を行なえばよいことはわかる。だが、1)の大量の専門分野の文書データの入手が大変困難である。特にパラレル・コーパスはなかなか入手できない。機械可読ファイルを望むならなおさらである。

〔6〕マルチリンガル専門用語

多言語の機械翻訳が望まれている。このためにはマルチリンガル専門用語の開発が急務である。日本の貿易先をみると英語圏ばかりではない。このためにマルチリンガル専門用語データ・バンクを次のようにして作る。



X国語を全ての人が理解することができないので英語を介して翻訳する。一部のものはX↔英、英↔日の用語がうまく一致するものも機械的に探し出すことも、自動的に出来る。

〔7〕おわりに

機械翻訳システムの評価は数文だけのテストで優秀の評価を行っている。しかし、専門分野の数とその適合度が重要である。だが、そこまで考慮した用語評価はまだまだ普及していない。評価の方法の確立と品質向上にむけた研究と実践が必要である。

〔8〕資料

- (1) ATLAS V4.0 for Windows 95/NT カタログ 1998年4月 (CZ10036)
- (2) 英日翻訳支援システム Version 2.0
たちまち翻訳 日立製作所カタログ (CD-646, 1998年1月)

〔9〕参考文献

- (1) 亀井 真一郎、平井 徳行他
商用機械翻訳ユーザー辞書の共通フォーマット設定に向けて 情報処理学会 第54回全国大会
4B-4 1998年3月