

# 科学技術用語の英日翻訳規則

6 B-2

田上 文俊<sup>†</sup> 竹田 正幸<sup>‡</sup> 松尾 文碩<sup>‡</sup><sup>†</sup>九州大学大学院工学研究科 <sup>‡</sup>九州大学工学部

## 1. まえがき

科学技術分野では、研究の進展に伴い次々と新たな専門用語がつくられる。したがって、科学技術文献の機械翻訳において、対訳辞書だけで対応するのは困難である。そこで、辞書に未登録の用語の英日自動翻訳法を開発するために、英日対訳辞書から翻訳規則の抽出を試みた。

## 2. 訳の抽出法の改善

著者らは、科学技術用語の英日対訳辞書に基づく単一語の訳語の抽出法を開発している<sup>1)</sup>。科学技術用語とその対訳のデータとして、市販の機械可読である日外アソシエーツ社 EB 科学技術用語大事典(以後、科学技術用語辞書という)を用いた。この対訳辞書の英語見出しとその訳の対の集合を  $T$  で表し、英語見出しが単一語であるものの集合を  $T_0$  で表す。 $T$  の異なり英単語数は 91,905、 $T_0$  の異なり英単語数は 61,885 である。

抽出の手順を説明する。まず最初に  $T_1 = T_0$  とする。そして、 $T$  と  $T_0$  をもとに単一語の新たな訳を抽出する。例えば、

(aerial pollution, 大気汚染)  $\in T$

(pollution, 汚染)  $\in T_0$

のような場合、(aerial, 大気) を抽出する。こうして抽出した新たな対訳を、 $T_1$  に加える。すべての抽出が終了したら、今度は  $T_2 = T_1$  とし、 $T$  と  $T_1$  をもとに抽出した対訳を  $T_2$  に加えていく。以下、新たに訳が抽出できなくなるまでこれを繰り返す。詳しい抽出の方法については文献 1 を参照されたい。

科学技術用語辞書  $T$  には、

(1,2,3-trihydroxybenzene, ピロガロール)

のように、化学物質名が多数含まれていた。これらは特殊な訳され方をする。このため、今回の抽出作業では、 $T$  から分野が化学、医学生物である対訳を除いた。これにより、 $T$  の異なり英単語数は 35,735、 $T_0$  の異なり英単語数は 20,866 となった。

前回の抽出で得られた抽出訳に、(circuit, 収回路) があつた。これは、まず

(acoustic absorption, 吸音)  $\in T$

(acoustic, 音)  $\in T_0$

より (absorption, 吸) を抽出し、これと

(absorption circuit, 吸収回路)  $\in T$

から抽出されている。

そこで、このような訳の抽出を抑制した。上記の例では、

(absorption, 吸収)  $\in T_0$

(absorption, 吸)  $\in T_1$

の二つの対訳のうちから、前者を優先して抽出に用いる。これにより、(circuit, 収回路) は抽出されなくなる。

新たな抽出の結果、 $T_5$  で語数の増加は収束した。 $T_5$  の異なり英単語数は 29,271 である。

## 3. 対応する英単語のない和訳部分

訳を合成するとき、日本語を補って訳す方がよい場合がある。 $T_0$  を使い、 $T$  中の見出し語に従って訳を合成してみる。例えば、

(ring, 環)  $\in T_0$

(network, ネットワーク)  $\in T_0$

のふたつの訳を合成すると「環ネットワーク」となるが、 $T$  には (ring network, 環状ネットワーク) とある。この場合は、「状」の部分に対応する英単語がない。このような和訳部分を集計した結果を表 1 に示す。

この表の上位のものについては、それぞれの出現状況を調べた。例えば「状」について、「環状ネットワーク」中の「環」のように、「状」の前方に位置する和訳を調査してみたところ、総数 81 のうち漢字の場合が 71 であり、そのうち一字の漢字である場合が 68 と多

English-Japanese Translation Rules for Technical Terms  
Fumitoshi Tanoue<sup>†</sup>, Masayuki Takeda<sup>‡</sup> and Fumihiko Matsuo<sup>‡</sup>

<sup>†</sup>Division of Engineering, Graduate School, Kyushu University 36, Hakozaki, Fukuoka, 812 Japan

<sup>‡</sup>Faculty of Engineering, Kyushu University 36, Hakozaki, Fukuoka, 812 Japan

表 1 対応する英単語がない和訳部分 (出現総数 3213)

| 頻度  | 和訳部 | 頻度 | 和訳部 | 頻度 | 和訳部 |
|-----|-----|----|-----|----|-----|
| 554 | の   | 37 | 部   | 14 | 水   |
| 319 | 用   | 35 | 動   | 14 | 圧   |
| 289 | ・   | 28 | 面   | 13 | 製   |
| 108 | 形   | 27 | 付   | 12 | 分   |
| 81  | 状   | 23 | 的   | 12 | 区   |
| 62  | 性   | 20 | 付き  | 12 | 化   |
| 60  | 響   | 19 | 色   | 11 | 電   |
| 49  | 型   | 18 | 波   | 11 | い   |
| 46  | 式   | 16 | 系   | 10 | 電圧  |

かった。このような調査結果は、抽出訳を使用する際に利用する。

また、「響」についての出現例を以下に示す。

(acoustic, 音)  $\in T_0$

(delay, 遅延)  $\in T_0$

(acoustic delay, 音響遅延)  $\in T$

このように「響」は、(acoustic, 音響)  $\notin T_0$  のために見られた例である。他に、「区」「電」も同様な理由で出現しており、これらは前の「状」などとは区別して考える。

#### 4. 対応する和訳部分のない英単語

今度は、対応する和訳部分のない英単語について、以下に例を示す。

(drill, ドリル)  $\in T_0$

(practice, 練習)  $\in T_0$

(drill and practice, ドリル練習)  $\in T$

和訳「ドリル練習」中には、and に対応する部分がない。このような英単語の集計結果を、表 2 に示す。of, to, in といった前置詞に混じって、electric, type などが出現している。これは、

(circuit, 回路)  $\in T_0$

(theory, 理論)  $\in T_0$

(electric circuit theory, 回路理論)  $\in T$

のような場合である。この場合は、(electric, 電気)  $\in T_0$  を用いて「電気回路理論」と訳すことができる。前置詞、接続詞が省略される場合についての調査結果は、抽出訳を使用する際に利用する。

#### 5. 英単語の語順とその和訳部分の語順が異なる場合

抽出訳を用いて訳を合成するときには、英単語の語順とその和訳部分の語順が異なる場合を考慮しなければならない。この、語順が変化する場合について調査した。その例を以下に示す。

表 2 対応する和訳部分がない英単語 (出現総数 6972)

| 頻度  | 英単語        | 頻度 | 英単語      |
|-----|------------|----|----------|
| 677 | of         | 44 | line     |
| 109 | electric   | 43 | plate    |
| 92  | to         | 40 | circuit  |
| 55  | type       | 38 | material |
| 55  | air        | 37 | magnetic |
| 50  | water      | 34 | out      |
| 48  | in         | 34 | molding  |
| 47  | electrical | 30 | color    |
| 45  | and        | 29 | mold     |

(fire proof, 耐火)  $\in T$

(fire, 火)  $\in T_0$

(proof, 耐)  $\in T_0$

$T$  で見出し語が英語 2 単語からなる対訳について、その訳が  $T_0$  の訳の接続となるものは 95,206 あり、そのうち語順が異なるものは 3,076 (全体の 3.23%) であった。

上の例は、2 字熟語が「～に (を) ～すること」という形で構成される場合である。「耐火」は、「火に耐えること」をさし、動詞のはたらきをもつ漢字が、目的語の前に位置する。このような動詞のはたらきをもつ漢字としては、ほかに「防」「吸」などがある。

同じ「耐」という漢字を含む 2 字熟語には、このような例もある。

(proof stress, 耐力)  $\in T$

(proof, 耐)  $\in T_0$

(stress, 力)  $\in T_1$

「耐力」という 2 字熟語は「～する～」という形をとっており、「耐える力」をさす。この場合は語順は変化せず、前者との区別が必要となる。

#### 6. むすび

訳の抽出方法に改良を加え、再び訳の抽出を行った。また、和訳部分に対応する英単語がない場合と、英単語に対応する和訳部分がない場合を調査した。さらに、英語 2 単語からなる見出し語と訳の対について、英単語の語順と和訳部分の語順が異なる場合を調査した。

なお、本研究は、一部文部省科学研究費補助金 (#07558162) の援助により行った。

#### 参考文献

- 1) 田上, 坂口, 竹田, 松尾: 科学技術用語の英日翻訳規則の抽出, 情報処理学会第 51 回全国大会講演論文集 3-81 (1995).