

2 単語科学技術用語の英日翻訳について

4 B-2

田上 文俊 竹田 正幸 松尾 文碩
九州大学大学院システム情報科学研究科

1. まえがき

科学技術文では、新しい技術用語が次々に生まれることから、その英日翻訳においては、辞書に未登録の科学技術用語の英日自動翻訳法の開発が必要である。この翻訳法として、単一語の訳語を組み合わせることによって行うことが考えられる。しかし、単一語の訳語は複数存在するので、それらの訳語の組合せが問題となる。正しい組合せを得るための制約条件は、長さや字種について検証済みである³⁾。本論文では、単一語の訳語として1漢字を用い、2単語句を2字熟語で翻訳する場合について調査した。

例えば、high temperatureという見出しに対し、「高温」という訳語が辞書¹⁾にある。highの訳語には、辞書中の訳語として「高、高位、高レベル」がある。temperatureの訳語には、辞書中の訳語として「温度」のみがある。high temperatureの訳語「高温」は、highの訳語「高」と、temperatureの訳語「温度」の先頭に出現する漢字「温」を組み合わせれば得られる。

2. 使用した辞書

科学技術用語の対訳のデータとして、EB科学技術用語大辞典¹⁾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。今回の作業では、 T から分野が化学、医学生物である対訳を除いており²⁾、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。

3. 使用する漢字の選出方法

ここでは、単一語の訳語として使用する1漢字を選出する方法について考察する。high temperatureの訳「高温」の場合は、highの訳語には1漢字訳「高」があるため、これを用いればよい。しかし、temperatureの訳語には、「温度」しかない。「温」と「度」のうち、どちらを選択すればよいのだろうか。

ここで、「温」という漢字をふくむ2漢字訳語をもつ英単語は、次の二つがある。

(spa, 温泉) $\in T_0$

(temperature, 温度) $\in T_0$

しかし、「度」という漢字をふくむ2漢字訳語をもつ英単語は、次のように64あった。

(angle, 角度) $\in T_0$

(brightness, 輝度) $\in T_0$

(concentration, 濃度) $\in T_0$

.....

.....

ここで「度」は、多くの英単語の2漢字訳語に出現している。このようなものは、英見出し語にとって情報量の少ない漢字と見なすことができる。そこで、このような漢字を、英単語の異なり語数によって集計した。この時の見出し語の異なり語数を、漢字冗長度数と呼ぶ。上位の結果を以下に記す。

漢字	冗長度数	漢字	冗長度数	漢字	冗長度数
分	94	板	60	子	53
石	88	座	59	数	51
水	74	気	56	口	51
線	73	出	55	定	50
化	71	形	55	電	48
度	64	流	54	地	48
合	63	動	54	金	48

ここで、漢字冗長度数が最も大きい「分」について見てみよう。

(clause, 分節) $\in T_0$

(increment, 増分) $\in T_0$

On English-Japanese Translation for Two-Word Phrase of Technical Terms

Fumitoshi Tanoue, Masayuki Takeda and Fumihiro Matsuo

Graduate School of Information Science and Electrical Engineering, Kyushu University, Hakozaki, Fukuoka, 812-81 Japan

(moisture, 水分) $\in T_0$

.....

.....

たしかに「分」は、見出し語との結び付きが弱いといえるが、次のように「分」を1漢字訳語として用いる場合もある。

(phase splitting, 分相) $\in T$

(phase, 位相) $\in T_0$

(splitting, 分流) $\in T_0$

(splitting, 分裂) $\in T_0$

phase splittingの訳語「分相」は、phaseの訳語「位相」の最終に出現する漢字「相」と、splittingの訳語「分流, 分裂」の先頭に出現する漢字「分」を組み合わせれば得られる。splittingの訳語中には、「分」という漢字が複数出現している。このように、同一見出し語の訳語中において、出現した回数が多い漢字を優先して使用することが考えられる。

4. 使用する漢字の選出実験

前述の方法にて、用いる1漢字を選出する実験を行った。本論文では、2単語句を2字熟語で翻訳する場合を扱っている。科学技術用語辞書より、そのような場合の対訳を集めて使用した。

(heat ray, 熱線) $\in T$

(heat, 熱量) $\in T_0$

(ray, 光線) $\in T_0$

上の例では、「熱線」は、heatの「熱」とrayの「線」からなると考えられる。このように、2単語句の2漢字訳語が、単一語の訳語中に出現する漢字を組み合わせ得られる場合は、全部で674であった。このうちから、前述の方法にて、使用すべき漢字をどれくらい選出できるかを調査した。

まず、訳語中での出現回数が最も多い漢字を選出し、同数の漢字がある場合に、漢字冗長度数の小さいものを優先して選択した。こうして漢字を選出した場合、305は辞書の訳語と一致した。

5. 英単語の語順とその和訳部分の語順が異なる場合

英単語の語順とその和訳部分の語順が異なる場合については、論文2で触れている。2字熟語について語順が入れ替わる場合には、次のようなものがあつた。

(sound reception, 受音) $\in T$

(sound, 音響) $\in T_0$

(reception, 受信) $\in T_0$

上の例では、動詞の働きをもつ漢字が先頭に來ている。

「受音」は、「音を受けること」を指し、「～に(を)～すること」という形で構成されている。この場合、動詞語義をもつ1漢字が、目的語の前に位置するため語順が入れ替わっている。そこでEDR日本語単語辞書を用い動詞語義をもつ784漢字を抽出し、このうちから有用と思われる707漢字を人手で選定した。

さらに、訳語の先頭語、最終語となる漢字が動詞語義をもつかどうかを調査し、語順が入れ替わる場合と比較した。実験には、前述の2単語句を2漢字訳語で翻訳している674の対訳を使用した。

実際に語順が入れ替わっている対訳は41である。このうちの36の対訳が、訳語の先頭語となる漢字が動詞語義をもっていた。

また、語順が入れ替わらない残りの633の対訳のうち、77の対訳では、訳語の最終語となる漢字が動詞語義をもっていた。このような場合について、出現した動詞語義をもつ漢字の頻度とその例を以下に記す。

漢字	出現頻度	出現例
流	19	気流, 熱流, 砂流, 定流, 水流, ……
動	4	浮動, 手動, 波動
組	3	床組
受	3	軸受, 側受
巻	3	鎖巻, 直巻, 波巻

前述の例と同じ「受」という漢字を含む2字熟語では、「軸受」などがある。この2字熟語は「軸を受けるもの」を指し、「～を～するもの」という形で構成されている。この場合は語順は変化しない。

他にも、2漢字が主従関係にある場合（「鶏鳴」「地震」など）は語順が入れ替わらないが、このような2字熟語の例は、この辞書中では見られなかった。

6. むすび

2単語句を2字熟語に翻訳する場合の、それぞれの単一語の訳語として用いる漢字の選出方法を提示し、自動翻訳実験を行った。また、英単語の語順と和訳部分の語順が入れ替わる場合について調査を進めた。

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) 日外アソシエーツ社：EB科学技術用語大辞典(1992)。
- 2) 田上, 竹田, 松尾：科学技術用語の英日翻訳規則, 情報処理学会第52回全国大会講演論文集 3-89 (1996)。
- 3) 田上, 竹田, 松尾：科学技術用語における2単語句の英日翻訳規則, 情報処理学会第53回全国大会講演論文集 2-63 (1996)。