

接頭・接尾辞を含む数量表現の翻訳方法

延原 由高 池原 悟 村上 仁一

鳥取大学大学院工学研究科

従来、数量表現の日英機械翻訳の研究では数詞、助数詞、名詞の組合せた表現のみが研究の対象とされていた。しかし、実際の文章では、接頭・接尾辞を含む表現もよく使用されている。そこで、本研究では、そのような表現を対象に翻訳規則の検討を行った。具体的には、新聞記事 10000 文の対訳コーパス中の数量表現から、それに含まれる接頭・接尾辞 40 種類を取り出し、使用頻度の高い 21 種類を対象に、助数詞と名詞の意味属性、数詞の精細度、数詞の桁数などに着目して翻訳規則を作成した。この規則を新聞記事、機能試験文集の数量表現の翻訳に適用した結果によれば、翻訳の正解率は、それぞれ 70%、66% となった。

Translation Rules for Numerical Expressions with Prefix and Suffix

Yositaka NOBUHARA Satoru IKEHARA Murakami JIN'ICHI

Tottori University

Conventional researches on Japanese to English machine translations have not dealt with the quantity expressions with prefix and suffix. This paper proposed translation rules for such expressions using the examples extracted from 10,000 sentences of bilingual corpus for newspaper. The rules are described by numerical classifiers, semantic attributes for nouns and effective values for numerals. As the results of having applied the rules to newspaper articles and a test sentence set for Japanese English machine translations, translation accuracy amounts to 70% and 66% for each.

1 はじめに

従来、数量表現の日英翻訳では助数詞とターゲット(対象とする名詞)の関係から翻訳規則を決定する研究 [1] や、数量詞とターゲットの意味関係から翻訳規則を決定する研究が行なわれている [2]。

しかし、これらの研究では、数詞、助数詞、名詞の組合せのみが対象とされており、接頭・接尾辞は研究の対象とされていない。

これに対して、接頭・接尾辞は新聞記事などにおいては数多く出現する。そこで本研究では、従来の翻訳規則、精細度、数詞の桁数、助数詞、名詞の意味属性などに着目した接頭・接尾辞の翻訳方法を提案する。

2 接頭・接尾辞の分類

2.1 接頭・接尾辞の定義

吉武 [3] らは数量表現を文節の一種と考え、助数詞の繰り返しや、区切りなどを考慮して定義した。しかし、接頭・接尾辞にあたる「数量接頭語」と「数

量限定語」については明確な定義は行われていない。そこで、本研究では数量詞に対する接頭・接尾辞を「数詞を含む接頭・接尾辞」と定め、ALT-JAWS の品詞体系で接頭・接尾辞に分類される単語をその対象とする。ALT-JAWS とは日英機械翻訳システム ALT-JE で使用されている日本語形態素解析プログラムで、本検討の対象とする接頭・接尾辞語は、それぞれ、品詞コード「6170(前置助数詞型)」の語と「6240(助数詞承接型)」の語である。本研究ではこの品詞コードを持つ単語が含まれる数量表現を対象に翻訳規則を検討する。以下、本研究の対象とする数量詞の定義と接頭・接尾辞の例を示す。

数量詞 = (数量接頭語)(符号)(数)(助数詞)(数量限定語)

数量接頭語	例：第、約、計、およそ
符号	例：+、-、±
数	例：整数、小数、分数、範囲数
数量限定語	例：強、以上、未満

表 1 接頭・接尾辞の定義

	名称	例
接頭辞	前置助数詞型 (6170)	約、平均など
接尾辞	助数詞承接型 (6240)	弱、以下など

2.2 接頭・接尾辞の出現頻度

まず、日本文中に出現する数量表現の種類と頻度を調べるため、辞書(和英辞典)と新聞記事(日本経済新聞)に使用された数量表現を抽出し、接頭・接尾辞を含む数量表現の割合を調べた。その結果を表2に示す。表2より、数量表現の使用される割合は、文当たりそれぞれ3.3%、13.1%であること、また、接頭・接尾辞を持つ数量表現は、文当たり、それぞれ、0.2%、5.9%で、特に新聞記事での使用震度が高いことが分かる。

表 2: 接頭・接尾辞の頻度

コーパスの種類	文数	数量詞の頻度	接頭・接尾辞の頻度
辞書	25000 文	832 件	51 件
新聞記事	10000 文	1315 件	589 件

2.3 接頭・接尾辞の分類

ここで、新聞記事 10000 文の対訳コーパス(以下、単に「対訳コーパス」と言う)中の数量表現(589 件)に使用された接頭・接尾辞(40 件)を意味の違いから表3に示すように6種類に分類する。

表 3: 接頭・接尾辞の分類

数の概略(8種)	約(138件)、前後(16件)など
数の上限(12種)	以上(47件)、以下(8件)など
数の増幅(2種)	増(34件)、減(8件)
年月の割合(8種)	月間(13件)、月産(14件)など
数の順序(1種)	第(100件)
数の性質(9種)	計(13件)、合計(7件)など

3 接頭・接尾辞の翻訳方法

対訳コーパスに使用されている40種類の接頭・接尾辞(用例数589件)のうち、本研究では使用頻度の高い19種類(用例数504件)を検討対象とする。

また、接頭・接尾辞の翻訳規則の記述では、ターゲットの名詞と接頭・接尾辞の意味的用法を表すため、「意味属性番号」を使用する。また、数詞で示される数量の細かさを表す尺度として、新たに「精細度」を定義する。ここで使用する「意味属性番号」と「精細度」は、以下の通りである。

「意味属性番号」

日本語語彙体系[5](岩波書店)で規定されたもので、約40万語の名詞の意味的用法を最大12段、2710の属性に分類した木構造の番号を表す。

「精細度」

例えば、「約1230個」と「約1000個」の表現を比べると、前者の示す数値が細かいのに対して後者の示す数量は大まかである。このような数量表現の表す数値の「細かさ」と「大まかさ」を表現するため、数詞の桁数に対する有効数字の割合を表す指標として「精細度」を以下の通り定義する。

$$\text{数の精細度} = \frac{\text{桁数} - 0 \text{ の桁の個数}}{\text{桁数}} \quad (1)$$

この値は、0~1までの範囲にあり、1に近付くほど細かさが増す。以下、精細度の計算例を示す。
例: 1234560000 → 6/10 = 0.6 (1に近付く程明確)
例: 1000000000 → 1/10 = 0.1 (0に近付く程曖昧)

3.1 「数の概略」

「数の概略」はおよその数を表している。対訳コーパスでは、「数の概略」の接頭・接尾辞には「約」(138件)、「前後」(16件)、「程度」(43件)、「およそ」(2件)、「ぐらい」(1件)、「くらい」(3件)、「ほど」(1件)、「とも」(1件)の8種類がある。このうち、ここでは頻度の高い「約」、「前後」、「程度」を検討対象とする。

ここでは、「数の概略」に対して翻訳規則を以下の3種類A、B、Cに分類し、検討する。また、訳文の生成では、規則A、B、Cの順に適用されるが、規則Aが適用された表現も場合によっては規則B、規則Cが適用される。同様に規則Bの適用され表現も場合によって規則Cが適用される。従って、最終的に翻訳方法が決定されるのは、規則Cの適用後である。

規則A: 従来の人間用の辞書や文法書から抽出した規則

規則B: 接頭・接尾辞の固有の規則

規則C: 複数の解を正解とする規則

3.1.1 規則A: 人間用従来の規則

規則Aは従来の従来の人間用の辞書や文法書から抽出した規則を正解率の高い順に適用するものである。「数の概略」は他の接頭・接尾辞と比べて使

用頻度が高く、さまざまな翻訳規則が考えられている [5,6,7]。その規則を対訳コーパスに適用し、正解率が高い規則を本研究で扱う規則として使用する。以下に本研究で適用した規則 a,b,c,d とその例を示す。

a:直前が"at"もしくは概数を表す動詞の場合訳は省略 [6]

例：私は経費を約 100 万円と見積もった。

→ I estimated the expenses at one million yen

b:数詞に前後の幅がある場合訳は"some"、または「数詞-数詞」+ "range" [6]

例：この金は約 20、30 キログラムです。

→ This gold weights some 20,30 kg.

→ This gold weights 20-30 kg ranges.

c:数詞が 10 の倍数以外、ターゲットが単数系の場合、訳は"some" 以外 [7]

例：駅まで約 25 キロの道のりだ。

→ It's a distance of × some 25km from the station.

例：このデータは約 1 ギガです。

→ This data is × some 1 giga.

d:数詞が助数詞の場合、訳は"around" 以外 [8]

例：これは約 5.5 ポンドです。

This weighs × around five pounds.

上記の点を考慮して図 1 に翻訳規則を示す。

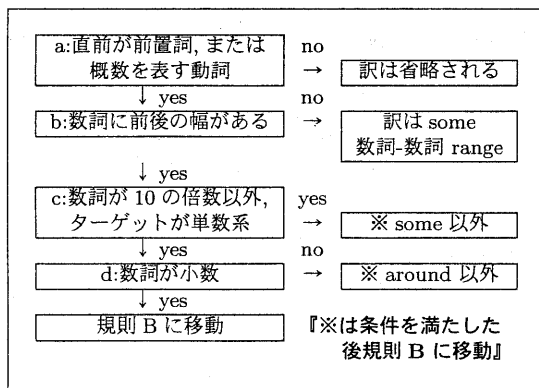


図:1 人間用従来の規則を用いた翻訳方法

3.1.2 規則 B：接頭・接尾辞の固有の規則

規則 A が「数の概略」を表す表現の意味的な分類の基づき翻訳方法を規定しているのに対して、規則 B は、それぞれの接頭・接尾辞（「前後」、「程度」）に対して個別に設けられた規則である。規則 A で適

用を受けた表現でも、規則 B の適用を受ける。以下に、「前後」、「程度」に対する翻訳規則を示す。

「前後」の翻訳方法

接尾辞「前後」の訳語として、"around" を与える。これは、対訳コーパスの例で、概略の数値を示す接頭・接尾辞「約」、「程度」の場合と比べて、「前後」では、"around" と訳される頻度が高いことによる。なお、この訳は、規則 C の適用の結果、見直されることがある。

「程度」の翻訳方法

「程度」の意味には以下の 2 種類がある [9]。

(1) 「程度の大小が物事の適否にかかわること」

例：砂糖を 10kg 程度 ならあげる。

(2) 「ちょうど適当と考えられる度合」

例：今年の予算は 1500 円程度 で済む。

対訳コーパスでは、(1) の意味の時は、「only」が使用されているが、(2) の意味の場合は、訳語として複数の候補が存在する。そこで、(1) の意味の場合のみ訳語を"only"とする。(2) の意味の場合は、訳語を生成せず、規則 C の適用を待つ。

なお、「約」は「数の概略」を表す接頭辞の中で最も使用頻度が高いが、その用法も多彩でユニークな訳語決定は難しいため、規則 B は設けず、複数の解を生成する規則 C を適用する。

3.1.3 規則 C：複数の解を正解とする規則

規則 C は訳を一意に決定せず、複数の訳を生成する規則である。訳語としては、5 種類の単語、"about"、"approximately"、"around"、"roughly"、"some" が使用される場合と省略される場合がある。以下、各訳語を選択するための規則を示す。

表 4：複数の解を正解とする翻訳規則

about	例外を除いてすべて正解
approximately	精細度が 0.3 以上の場合
around	数詞が 100 の位以下の場合 規則 B より「前後」のみ全て正解
roughly	精細度が 0.17 以下の場合
some	数詞が「万」以上の場合
省略される	直前に前置詞が来る場合

3.2 「数の上限」

「数の上限」は数詞の上限を表している。対訳コーパスでは、「数の上限」の接頭・接尾辞には「以上」(47件)、「以下」(8件)、「以内」(3件)、「強」(9件)、「弱」(7件)、「高」(2件)、「安」(2件)、「未満」(3件)、「最大」(5件)、「最大限」(1件)、「最高」(2件)、「最低」(1件)の12種類がある。このうち、ここでは頻度の高い「以上」、「以下」、「以内」、「強」、「弱」を研究対象とする。

3.2.1 「以上」の翻訳方法

例えば日本語で「10以上」と言うときは10が含まれるが、“more than”、“over ten”には10は含まれない。厳密さが必要な場合は、“ten or more”、“ten or over”と訳されるが、厳密さが重要でないときは、“more than ten”、“over ten”を使うことが可能である [10]。

- 例： 彼女はケーキを10個以上食べた。
- She ate ten cakes or more.
 - She ate more than nine cakes.
 - She ate over nine cakes.

上記の例で示されるように、“more than”、“over”をつかって、接尾辞「以上」の意味を厳密に訳す場合は、指定の数から1少ない数値を使用する必要がある。しかし、対訳コーパスでは、1程度の数の影響が表れにくい桁数の大きい数値の場合は、原文の数値に対して“more than”、“over”が使用されている。そこで、対訳コーパスから“more than”、“over”が使用された表現の数値を調べた結果に基づき、「精細度」0.65以下の場合に限り、これらの訳語を使用する。なお、訳語“over”、“above”は、助数詞が度数を表す場合のみ適用する。ここで、「度数を表す助数詞」とは数の単位を表す助数詞で、対訳コーパスにおいては「円」、「ドル」、「Kg」、「%」などである。以上の点を考慮して表5に「以上」の翻訳方法を示す。

表5: 「以上」の翻訳方法

精細度が 0.65以上	助数詞が 度数を表す	英訳
○	○ ×	or over or more
×	○	over above more than
×	×	more than

3.2.2 「以下」の翻訳方法

接尾辞「以下」の訳には“less than”、“under”、“below”がある。これらは、接尾辞「以上」に対する訳語、“more than”、“over”、“above”の反対語であり、「以上」の場合と同様の翻訳規則が適用できる。以上の点を考慮して表6に「以下」の翻訳方法を示す。

表6: 「以下」の翻訳方法

精細度が 0.65以上	助数詞が 度数を表す	英訳
○	○ ×	or under or less
×	○	under below less than
×	×	less than

3.2.3 「以内」の翻訳方法

対訳コーパスでは、接尾辞「以内」は主に“less than”、“or less”、「within」の3種類に翻訳される。また、「以内」は「以下」と同様、数詞の表す数値を含むため、同様の精細度0.65以下の場合に限り、“less than”、“or less”の訳語を使用する。なお、“within”は数詞自身の意味を含むため、数詞の「精細度」と無関係に適用可能とする。以下に例を示す。

※精細度が0.65未満

例:彼は出費を1000円以内におさえた。(精細度0.25)

→ He kept expenses to within a thousand yen
less than

※精細度が0.65以上

例:次の記事を15語以内書きなさい。(精細度1.00)

→ Write the following article in within 15 words.
15 words or less.

以上の点を考慮して、図2の通り「以内」の翻訳規則を決める。

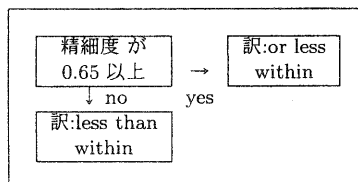


図2: 「以内」の翻訳方法

3.2.4 「強」の翻訳方法

対訳コーパスでは、接尾辞「強」は主に“more than”、“over”の2種類に翻訳される。また、“over”

は桁数が大きい場合に使われることが多い [7]。そこで、翻訳の基準とする数詞を「一億以上」と定め、数詞が「一億以上」なら、「強」の訳は“over”、以下なら“more than”とする。以下に例を示す。

例：10 億円強 → over one billion yen
 例：1000 円強 → more than thousand yen

以上の点を考慮して以下の図 3 に「強」の翻訳方法を示す。

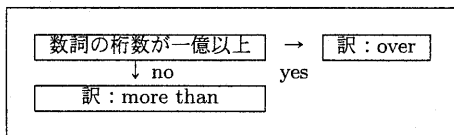


図 3: 「強」の翻訳方法

3.2.5 「弱」の翻訳方法

対訳コーパスでは、接尾辞「弱」は主に“less than”、“under”の2種類に翻訳される。また、“under”は「強」同様に桁数の多い場合に適用される。そこで翻訳の基準とする数詞を「一億以上」とする。以上の点を考慮して図 4 の通り、「弱」の翻訳規則を決める。

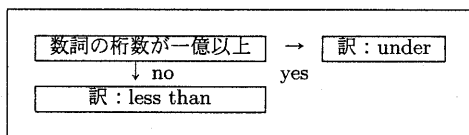


図 4: 「弱」の翻訳方法

3.3 「数の増幅」

「数の増幅」は数の増加、減少を表している。対訳コーパスを調べた結果、「数の増幅」の接頭・接尾辞には「増」(34 件)、「減」(8 件)がある。このうち、ここでは頻度の高い「増」を検討対象とする。

3.3.1 「増」の翻訳方法

対訳コーパスでは、接尾辞「増」は主に“increase”、“increase+of”、“over”の3種類に翻訳される。また、数量表現が「数量詞+の+名詞」(Q の N 型)や「名詞+数量詞」(QN 型)の場合、“increase”、“increase+of”が使用されるのに対して、それ以外の方では、“over”が使用される。以下に例を示す。

例 1: 数量表現が Q の N 型、QN 型以外
 80%増 → 80% over

例 2: 数量表現が Q の N 型、もしくは QN 型の場合

前年度比 80% → 80% increase of the previous year
 the previous year, 80% increase

以上の点を考慮して図 5 の通り「増」の翻訳規則を決める。

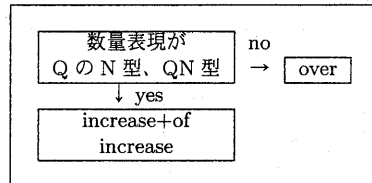


図 5: 「増」の翻訳方法

3.4 「数の性質」

「数の性質」は数量の性質を表している。対訳コーパスでは、「数の性質」の接頭・接尾辞には「平均」(9 件)、「合計」(7 件)「計」(13 件)、「同」(5 件)、「台」(12 件)、「当たり」(12 件)、「中」(1 件)、「分」(1 件)、「余り」(6 件)の 9 種類がある。このうち、ここでは、頻度が高くて代表的な「平均」、「合計」、「計」の 3 種類を検討する。

3.4.1 「平均」の翻訳方法

対訳コーパスでは接頭辞「平均」は主に“average of”と“in average”の2種類に翻訳される。また、「平均」は通常“average of”と訳されるが、接頭辞「当たり」と名詞句になる場合“in average”と訳される。以下に例を示す。

※数量詞に接辞「当たり」が含まれる場合

例：一キログラム当たり平均一千五百円
 → 1500 Yens per Kg in average

※数量詞に接辞「当たり」が含まれない場合

例：平均 5.2% → average of 5.2 percent

以上の点を考慮して、図 6 の通り「平均」の翻訳規則を決める。

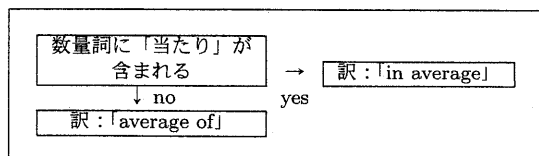


図 6: 「平均」の翻訳方法

3.4.2 「計」、「合計」の翻訳方法

対訳コーパスを調べた結果、接頭辞「計」、「合計」の訳には主に“a total of”、“totaling”、“訳が省略される”の3種類がある。これらの訳には以下に示す規則が適用できる。

- 1: 動詞が“include”、“occupy”、“amount”のように数の合計を含む意味が含まれている場合訳は省略される。
- 2: 「計」の訳の後に前置詞が来る場合、“a total of”は前置詞(of)+前置詞となり、文が冗長になるため訳は“totaling”が適している。

以上の点を考慮して、接頭辞「計」、「合計」の翻訳方法を図7に示す。

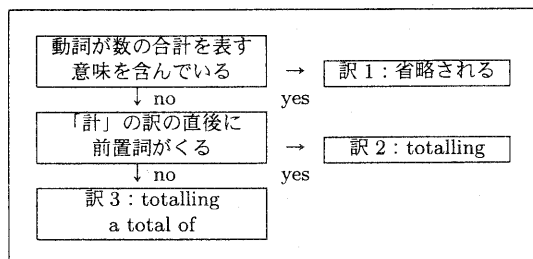


図7: 「計」、「合計」の翻訳方法

以下に「訳1」、「訳2」、「訳3」の例を示す。

「訳1」 - 「求書の金額は合計5万円に達した。」

The bills amounted fifty thousand yen.

「訳2」 - 「彼の収入はおよそ合計20万円です。」

His income is totaling about 200000 yen.

「訳3」 - 「出席者は合計200人だった。」

The attendance was totaling a total of two hundred.

3.5 「数の順序」

3.5.1 「第」の翻訳方法

対訳コーパスでは、「第」は序数詞に訳されている。しかし、助数詞が「号」、「番」などを示す場合、「No+数詞」、もしくは「序数詞」に訳される。また、ターゲットの意味属性番号が「憲法(1163)」以下に属する場合「Article+数詞+of+数詞」に訳される。以下に例を示す。

例1: 第二ラウンド → the second round

例2: 第30号 → No.30

例3: 憲法第9条 → Article 9 of the Constitution

以上の点を考慮して以下の表に「第」の翻訳方法を示す。

表6: 「第」の翻訳方法

助数詞が番号	名詞の意味属性番号(1163)	翻訳結果
○	×	No+数詞 序数詞
×	○	Article+数詞+of+名詞
×	×	序数詞

3.6 「年月の割合」

「年月の割合」は接辞が数詞の時間の割合を表している。対訳コーパスを調べた結果「年月の割合」の接頭・接尾辞には「月間」(13件)、「月産」(14件)、「月」(3件)、「年間」(19件)、「年産」(7件)、「年」(14件)、「年率」(9件)、「日量」(7件)の8種類がある。このうちここでは頻度が高い「月間」、「月産」、「月」、「年間」、「年産」、「年」を検討対象とする。

3.6.1 「年間」、「年産」、「年」の翻訳方法

対訳コーパスでは、接頭辞「年間」、「年産」、「年」は主に“per year”、“a year”、“yearly”、“annual”の4種類に翻訳されている。“per”は商業的意味合いをもつときに使用されることが指摘されている[10]。対訳コーパスでは、確かに「per」は、助数詞が「ドル」、「円」などの金額の意味を持つ場合、もしくは数量詞と共に名詞が「商業的意味合い」を持つ場合に適用されていることが確認できる。そこで、「商業的意味合い」を持つ名詞を意味属性番号「1867(取引引き)」と「1920(労働)」以下に属する名詞と規定し、これらの名詞をターゲットとする数量表現の場合、「年間」、「年産」、「年」の訳語として“per”を使用する。それ以外の名詞をターゲットとする場合は、“yearly”と“annual”を使用する。以上の点を考慮して「年間」、「年産」、「年」の翻訳規則を図8に示す。

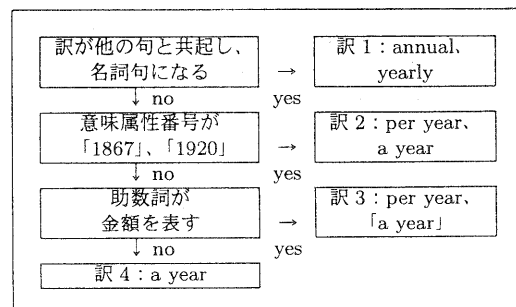


図8: 「年間」、「年産」、「年」の翻訳方法

以下に「訳1」、「訳2」、「訳3」、「訳4」の例を示す。

「訳1」 - 「私はこの分野の売上を年間1億円まで引き上げた。」

→ I increased the the annual proceeds
to one hundred million yen on this field.
the yearly proceeds

「訳2」 - 「年間三十万個の量産」

mass-production of 300000 units per year
a year

「訳3」 - 「会費は年間三千元です。」

Membership fees are 3000 yens per year
a year

「訳4」 - 「生産レベルは年間二万台になるだろう。」

The production level will be raised to 20000 units
a year

3.6.2 「月間」、「月産」、「月」の翻訳方法

対訳コーパスを調べた結果「月間」、「月産」、「月」の訳には”monthly”、“a month”、“per month”がある。また、“monthly”、“a month”、“per month”は”annual”、“yearly”、“a year”、“per year”と類義語のため、「年間」、「年産」、「年」と同様の規則が適用できる。以上の点を考慮して「月間」、「月産」、「月」の翻訳規則を図9示す。

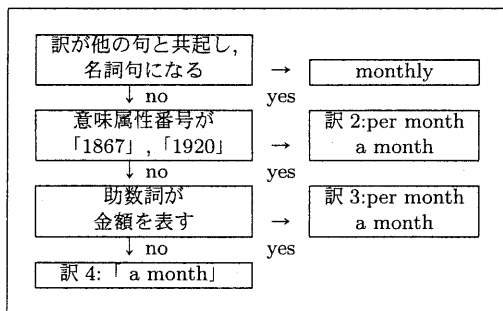


図9: 「月間」、「月産」、「月」の翻訳方法

4 評価実験

4.1 実験

3章で作成した翻訳方法の精度を新聞記事、機能試験文集の2種類の対訳コーパスを使用して評価する。機能試験文集は、日英機械翻訳の機能をテストするために編集された約6200文の対訳コーパスである。また、本研究で提案した翻訳規則の有効性を評価するため、頻度が最も高い翻訳を正解とす

る翻訳方法の結果(以降「デフォルト値」で示す)とも比較する。

- 1:新聞記事に含まれる数量表現(クローズドテスト)
- 2:機能試験文集に含まれる数量表現(オープンテスト)

なお、新聞記事を対象とする評価実験は、翻訳規則の検討で使用したデータと同一のデータを対象とするもので、クローズドテストであるのに対して、機能試験文集を対象とする評価実験は、オープンテストである。評価方法と評価を以下に記す。

$$\text{正解率} = \frac{\text{正しい翻訳候補}}{\text{全データ}}$$

表3:評価実験

	本研究の翻訳方法	デフォルト値
	正解率	正解率
新聞記事	70%	49%
頻度(クローズド)	(354/504)	(247/504)
機能試験文集	66%	61%
頻度(オープン)	(45/69)	(42/69)

4.2 実験結果

評価実験の結果を以下に示す。

- ・新聞記事の場合、提案した規則の正解率が70%であるのに対して、デフォルトの正解率は49%で、本論文の規則では、正解率がデフォルト値より21%上昇した。
- ・機能試験文集では、提案規則の正解率66%に対して、デフォルトの正解率は61%で、正解率の向上は、5%に止まっている。

上記の結果より、本研究で考案した翻訳方法は、新聞記事の翻訳において、効果的であることが分かる

4.3 成功例・失敗例

今回の実験で、翻訳に失敗した用例は、(1)接頭・接尾辞に対して、模範訳と異なる訳語が生成された場合(2)規則による翻訳では接頭接尾辞の訳語を生成したのに対して、模範訳では、翻訳不要として訳語が省略されたものの2種類に分けられる。特に、後者の場合は文のニュアンスに依存する傾向が強い。そのため接頭・接尾辞を翻訳規則を決めるうえで困難な問題の一つなると考えられる。以下に成功例と失敗例を示す。

(1) 成功例

100億円以上 → more than 10 billion yen

(2) 失敗例 1: 翻訳結果が異なる場合

5000 円以下の罰金 → × less than 5000 yen of fine
○ a maximum fine of 5000 yen

(3) 失敗例 2: 翻訳結果が省略される場合

一千個強 → × more than 1800000 units
○ 180000 units

(×: 本研究の翻訳方法、○: 対訳コーパス)

5 考察

5.1 評価結果について

オープンテスト、クローズドテストを行なった結果、本研究で考案した翻訳方法の正解率はそれぞれ、70%、66%であった。また、デフォルト値はそれぞれ 49%、61%であった。これより、本研究の正解率はデフォルト値より上昇することが解った。

実験結果によれば、新聞記事の場合は、デフォルト値に比べて翻訳精度が向上したのに対して、翻訳機能試験文の場合の翻訳精度の向上効果は小さい。これは、オープンテストとクローズドテストの違いにもよるが、新聞記事では、接頭接尾辞の用法がかなり限定されていて、正解率 (75%) の高い「数の概略」の出現頻度 (39%) が高いのに対して、機能試験文の数量表現では、接頭・接尾辞が多様な意味に使用されていることが原因と見られる。

また、新聞記事の場合も、翻訳の精度は必ずしも満足できるレベルにはないが、これは、本検討による規則が、接頭・接尾辞に対して訳語を生成することを前提としているのに対して、実際の翻訳では、訳語が省略される場合がかなり存在することが原因で、今後、省略の可能な条件についても検討を行う必要がある。

6 あとがき

本研究では新聞記事 10000 文の対訳コーパスに現れた数量表現のうち、接頭・接尾辞をもつ表現を対象に翻訳規則の作成を試みた。具体的には、助数詞、名詞の意味的用法を区別するため、すでに提案されている「単語意味属性」を使用するとともに、新たに、数詞の有効桁数に着目した「精細度」の基準を設けて、本格規則を作成した。この規則を新聞記事 (クローズドテスト) と機能試験文集 (オープンテスト) に適用した結果によれば、新聞記事の場合 70%、機能試験文集の場合 66% の精度で、接頭・接尾辞を含む数量表現の翻訳が出来ることが分かった。

本検討では、接頭・接尾辞に対して訳語を生成することを前提として翻訳規則を作成したが、実際の翻訳では、訳語が省略される場合がかなり存在する。今後、省略の可能な条件についても検討を行う必要がある。

参考文献

- [1] "When and How to Disambiguate?" PROCEEDING OF THEMIDDING-96 SEMINARL(1996)
- [2] 旭岡 他: "日本語数量表現の意味処理に付いて"、自然言語処理研究会 Vol90, No.78. P129-136(1990)
- [3] 吉武 他: "日本語数量表現とその英語への機械翻訳に関する一考察"
- [4] 宮崎、白井、池原: "言語処理過程説に基づく日本語品詞の体系化とその効用"、自然言語処理論文誌, Vol.2, No.1, pp.12-21(1995.1)
- [5] 池原、宮崎、白井 他: "日本語語彙体系"、岩波書店 (1997)
- [6] "WRITER'S GUIDE TO ENGLISH USAGE"
- [7] "A DICTIONARY OF ENGLISH WORD GRAMMAR ON ADJECTIVE AND ADVERBS"
- [8] "前置詞活用辞典"
- [9] 新村 出: "広辞苑 第五版"、岩波書店 (1998)
- [10] 長谷川、アラン、橋本: "プロシード和英辞典"(1988)
- [11] 延原、池原、村上: "接頭・接尾辞を含む翻訳方法" 第 6 回自然言語処理学会 p260-263(2000)