日英語の再帰形におけるアラインメント問題

清水 眞　　宇津呂 武仁
九州工業大学工学部　奈良先端科学技術大学院大学
伝 康晴　　中村 順一
奈良先端科学技術大学院大学　九州工業大学情報工学部

1．はじめに

二言語間の翻訳の際、訳語の対応が一対一ではないというアラインメント問題がしばしば起こるが、日英語の翻訳も例外ではない。例えば、「彼」、「彼自身」、「先生」、およびhe、himself、the teacher等の日英語の指示表現である。日英語で、それぞれ代名詞、再帰形、名詞句と呼ばれることが多い。従来の研究では、日本語の代名詞、再帰形、名詞句は、ほとんどの場合、英語の代名詞、再帰形、名詞句に対応することが、明示的にであれ、暗黙のうちにであれ、前提とされてきた。ところが、次のような問題が存在する。英語の再帰形（himself、myself等）を日本語に翻訳する場合を例に取ると、これまでは自動的に「自分」と訳すとされていた。しかし、日本語の電子化コーパスを見ると、「自身」、「自体」、「自己」等に訳す方が自然である事例も多いことに気がつく。また、John killed himselfの翻訳である「ジョンは自殺した」に見られるように、形態素「自」を接頭語とする名詞を含むサ変動詞に対応することも頻繁にある。by itselfやin itself等は、「ひとりでに」という副詞句に翻訳すべきであるという点は、意外なことに指摘している研究がほとんどない。一部の研究に、代名詞、あるいはゼロ代名詞に翻訳する方がいい場合があるという主張がなされてはいるものの、部分的な指摘にとどまり、その例文は不自然である[1]。

本研究の目的は、英語の再帰形の日本語への翻訳例を注意深く観察し、充分な検証を行い、それぞれを分類し、両者の対応を取ることである。特定の統語理論に依存せず、言語事実を指摘するという立場に留意している。電子化コーパスから抽出した実際の用例を用い、英語と日本語の対応をより現実に即した形で、分析、考察する。これまでの言語理論は、理論内の整合性のみに注意を払い、実際の言語運用を無視、あるいは軽視するきらいがあった。本研究では、自然な用法のみデータとして用いることを強調したい。そのため、用例を他の研究に用いることが可能である。

2．対訳テキスト

我々は、データとして対訳テキストを用いた。具体的には、計算機上で利用可能な和英辞典である、講談社学術文庫の和英辞典[2]から、約40,000件の対訳例を取り出し、対訳コーパス[3][4]としたものである。

対訳テキストは，自然言語処理のためのさまざまな言語的知識や 機械翻訳のためのさまざまな翻訳知識を取り出すための情報源として有用である。対訳コーパスからの翻訳規則学習の手法を用いれば、多大なコストを必要とせずに、首尾一貫性のとれた大量の翻訳規則を獲得することが可能であるからである。これまで翻訳規則が記述されていない専門的分野のテキストを翻訳するというタスクに対しても、対訳コーパスからの翻訳規則学習の手法を用いれば、特定の分野に特化した翻訳規則を低コストで獲得することが可能である。また、人間の内省に基づく規則記述ではとらえきれないような言語現象に対しても、翻訳規則の形式で対訳コーパスから取り出すことが可能である。

対訳テキストを解析し、そこから言語的知識や翻訳知識を取り出すための主要なアプローチの一つとして、統計的アプローチによる方法がある。対訳コーパスベース以外のアプローチには、伝統的なものとして、翻訳規則を人手で記述する、規則ベースのアプローチがある。Muプロジェクトが一例である。このアプローチの問題点は、人手で大量の規則を記述するため、多大なコスト、労力が必要なことである。また、大量の規則間の首尾一貫性を保持していくことも容易ではない。規則数が数百個以上になってくると、細かな修正に対してシステム全体の挙動を予測することが困難になってくる。

3．データ分析

3.0 結果

約40,000の対訳例から、英語の再帰形269例を抽出し、日本語訳を分析、分類した。パーセントは、小数点第三位を四捨五入した。

1 い型＝再帰形　22例=8.18%
- a. 自分　19例=7.06%
- b. 自ら　2例=0.74%
- c. 自体　1例=0.37%

2 （ろ）型＝代名詞　7例=2.60%
- a. 我　5例=1.86%

b. 僕にだって　1例=0.37%
c. きみ　1例=0.37%

3 (は) 型＝名詞句　66例=24.5%
a. 身体、身体の一部　18例=6.69%
b. 思考、言説、心理　16例=5.95%
c. 健康状態、疲労度　8例=2.97%
d. 容姿　2例=0.74%
e. 才能、能力　3例=1.12%
f. 言動、行動、行動様式19例=7.06%

4 (に) 型＝接頭辞を含む名詞句12例=4.46%
a.自+　10例=3.72%
b.独+　2例=0.74%

5. (ほ) 型＝ゼロ代名詞163例=60.6%
a. 自動詞　22例=8.18%
b. (接頭辞+名詞)+サ変動詞28例=10.40%
c. 目的語+他動詞72例=26.77%
d. その他　41例=15.24%

3.1　(い) 型＝再帰形

(1) a. You must advertise yourself more if you want to get the position.
その地位につきたいのなら君はもっと自分を売り込まなきゃだめだ
b. I am sorry I cannot express myself to my heart's content.
私は自分の思っていることを十分に言い表わせないのが残念だ
c. She is looking at herself in the mirror.
彼女は鏡の中の自分の姿をながめている
d. He justified himself by showing evidence.
彼は証拠を見せて自分の正しいことを証明した
e. He thinks himself somebody.
彼は自分で偉いと思っている
f. I flatter myself that I am equal to the task.
自分で言うのもおこがましいがそんな仕事ならばやってみせる
g. I myself don't undertand why I failed,
どうして失敗したのか自分でもわからない
(2) a. The plan itself is absurd.
その計画自体おかしい
b. He says he holds himself to blame.
彼は自らその罪を負うと言っている

従来の日本語の再帰形に関する論文では、「太郎は自分を愛している」のような例文が頻繁に用いられていた。英語の再帰形が完全他動詞の目的語で、それに対応する日本語文では再帰形が単独で用いられるような例である。このタイプはこのコーパス中には1例しか存在しない。(1a)である。他の例では、(1b-d)のように日本語文で再帰形が「姿」、「思っていること」、「正しいこと」等の名詞句により補われていたり、(1e, f)のように英語の再帰形が補語や目的語を伴ったり、英語の再帰形が副詞（この場合はいわゆる同格）の位置にあったりする。

これまでの研究ではあまり指摘されていないが、日本語には「自分」以外の再帰形が存在する。「自ら」、「自体」、「自身」、「自己」、「己」等である[5]。「自分」以外は再帰形ではないとする議論が予測できるが、その場合、(1)と(2)の分布が極めて似ているという事実を説明しなければならない。我々は、「自分」以外も再帰形であるという立場をとる。統語的な振る舞いという点を考えると、(2b)は特に興味深い。英語では目的語、日本語では副詞と品詞が異なるからである。

3.2　(ろ) 型＝代名詞

日本語に代名詞という統語範疇を認めるかということはさかんに議論されている。我々は認めないという立場をとるが、ここでは便宜上「代名詞」という用語を用いる。

(3) a. Look at yourself in a mirror.
鏡にきみの姿を映してごらんなさい
b. I am ashamed of my self for having done such a thing.

こんなことをしてわれながら浅ましい
c. I came to myself soon.
間もなく我に返った
d. He was beside himself with anger.
彼は怒りにわれを忘れた
e. He makes himself despised by what he does.
彼はみずからわが身を卑しめている
f. I myself can do that much.
そのくらいのことは僕にだってできる

日本語には代名詞はないという立場をとる我々であるが、(3a)の「きみ」、(3b)の「われ」が代名詞であるという主張は、まだしも理解できる気がする。しかし、(3c)の「我」が代名詞であると言うなら、これには異議を唱えなければならない。(3d)からわかるように、多分に慣用句的であるからである。日本語に代名詞はないという有力な証拠であるかもしれない。(3e)は、英語の再帰形が、日本語のいわゆる代名詞+名詞に対応する例である。(3f)では、英語の再帰形が主語と同格であり、日本語訳で対応するものが格助詞「にだって」として実現している。

3.3　(は) 型＝名詞句

(4) a. He was so tired that he dragged himself along.
彼は疲れた手足を引きずって歩いた
b. I cannot express myself fully in English.
英語では思っていることが言い尽くせない
c. Please take good care of yourself in this cold season.
体には十分気をつけてください。
d. He wears such strange clothes in order to make himself look eccentric.
彼は奇をてらってあんな格好をしている
e. You must improve yourself in English.
君は英語の実力をつけなければいけない
f. He devoted himself to the improvement of society.
彼は社会を改善することに一生をささげた

(4a-c)に関しては、「自分の (が)」、「彼の (が)」等が削除されているという議論が予測される。(1b-d)や(3a,2e)に示したとおり、確かにそのような例も存在する。しかし、その数は圧倒的に少ない。また、我々の言語直感に照らせば、(4a-c)に「自分の (が)」、「彼の (が)」等を挿入すれば、かえって不自然になる。よって、我々は、(は) 型は (い) 型や (ろ) 型を削除変形して得られたという主張を退ける。

はたしてこのように多くのタイプに分類しなければいけないのだろうかという疑問も湧くかもしれない。例えば、(4f)の「一生」を「自分」に書き換えれば「言動、行動、行動様式」タイプは必要なくなるのではないかという議論である。事実、(5)のような例が存在する。比喩的な表現ということを踏まえれば、「物理的な身体全体あるいは身体の一部」タイプに統一できる。

(5) He sacrificed himself for his country.
彼は身命を国に捧げた
b. He threw himself into the vortex of political strife.
彼は政争の渦中に身を投じた

しかし、我々は、「一生」を「自分」に書き換えられるという主張は退ける。意味的に似ているという、根拠が薄弱な理由で安易な書き換えはできない。「物理的な身体」対「比喩的な身体」という議論は、非常に興味深いが、比喩の理論がまだ確立していない現時点では、異なるタイプを統一するよりも、違うタイプとする方がより現実的、実際的であると考える。

3.1でも指摘したが、英語の再帰形が翻訳される際、日本語訳では、しばしば動詞+「こと」/「もの」に対応している。英語から日本語に翻訳される際に、統語的な範疇が変換していることに注目したい。

この他にも名詞句に翻訳されたと考えられる例が存在するのだが、後で議論する理由により、(ほ) 型に分類した。

3.4　(に) 型＝接頭辞を含む名詞句

(6) a. He gave himself up to the police, conscience stricken.
良心の呵責にたえかねて警察に自首した
b. He decided for himself without consulting his colleagues' opinions.
彼は同僚たちの意見を聞かないで独断で事を処理した

(7)は「自+」と分類したが、「自慢+」というタイプを設定した方がいいのかもしれない。意味的により限定されるからである。

(7) His proud long talk of himself spoiled the pleasure of the company.
彼の長と語る自慢話にはみんな興冷めした

3.5 (ほ) 型＝ゼロ代名詞

我々はゼロ代名詞という統語範疇を認めないという立場をとるが、ここでは便宜上「ゼロ代名詞」という用語を用いる。英語の再帰形が日本語に翻訳される際、最も頻繁な例である。

(8) a. He concealed himself behind the gate.
彼は門の陰に隠れた
b. A new doctor established himself on this street.
新しい医者がこの通りで開業した
c. I could talk myself out of the difficulty.
私はこの苦境を言い抜けられた
d. I could never bring myself to say such a thing.
とてもそんなことを言える義理じゃない

(8a)では、英語の再帰形が、再帰形はおろか名詞句としてさえ実現していないことに注目されたい。

(8b,c)では英語の再帰形が名詞句を含んだ動詞句に翻訳されている、よってゼロ代名詞ではない、という議論が予想される。3.3で、（は）型以外にも名詞句に翻訳されたと考えられる例が存在するのだが、（ほ）型に分類すると述べた。（は）型の名詞句が、身体、思考、行動等、再帰形で現される人物の属性に深く関与しているに比し、（ほ）型の名詞句はその関与が断片的である。例えば、(9)である。

(9) a. He devoted himself to the establishment of world peace.
彼は世界平和の達成に献身した
b. He has devoted himself to the study of cancer for a long time.
彼は十年一日のごとくがんの研究に専念している

一見、「言動、行動、行動様式」タイプの(4f)に類似しているように思えるかもしれない。しかし、(4f)と異なり、名詞句が英語の動詞の意味まで含んでしまっている。あるいは、（に）型である接頭辞を含む名詞句(6)に類似しているように思えるかもしれない。これもまた、英語の再帰形himselfの持つ意味の一部である「自+」「独+」を含む(6)とは異なり、接頭辞が実現するのは英語の動詞の意味である。よって、（に）型とは別の（ほ）型に分類した。

(8d)には単一のタイプしかあげていないが、その他にはこれ以外のタイプもある。

4 おわりに

以上、英語の再帰形の日本語訳を分析、分類した。日本語において再帰形に対応するものは8%弱で、残りも様々なタイプに別れるという点を確認しておきたい。しかしながら、英語の再帰形と日本語訳の間の対応が全くアトランダムであるというわけではなく、そこにはある種のパターンが存在するように思われる。例えば、日本語で再帰形に翻訳されるものは、英語で心理的な動詞、話者指示的な動詞と共起している傾向が強いようである[6]。日本語でゼロ代名詞+自動詞に翻訳されるものは、英語の再帰形が用いられる理由が主に統語的な制約である場合が多いように思える。共起関係を中心に研究をすすめたい。

今回の発表では、英語の再帰形の日本語訳のみを論じたが、我々はすでに日本語の再帰形の英語訳の分析にとりかかっている。ここにおいてはまた違った対応が存在している。問題は、日英語の再帰形にそれぞれ複数の訳語が対応するというのみにとどまらない。その対応が一対一ではないという点を認識すべきである。さらに、全体像を複雑にする要因は、この問題が、代名詞、名詞句、ゼロ代名詞すべてあてはまるということである。今後の課題としたい。

参考文献

[1] Kameyama, M.: Subjectivity/logophoric bound anaphor zibun, in Drogo, V., Mishra, V., & Testen, D. (eds.) Papers from the 20th Regional Meeting, Chicago Linguistic Society, Chicago: University of Chicago Press, pp. 228-38, (1984).
[2] 清水護, 成田: 和英辞典, 講談社学術文庫, (1979).
[3] 宇津呂, 松本: 対訳辞書および統計情報を用いた二言語対訳テキスト照合, コンピュータソフトウェア, Vol. 12, No. 5 (1995).
[4] 北村, 松本: 対訳コーパスを利用した翻訳知識の自動獲得, 情報処理学会論文誌,Vol. 37, No. 6, pp. 1030-1040, (1996).
[5] Aniya, S.: A Categorical Approach to Fundamental Problems in Japanese Syntax, Ph. D. Dissertation, University of Washington, (1987).
[6] 清水眞: 談話表示理論による再帰形の記述, 九州工業大学研究報告(人文･社会科学), 第38号, pp. 35-57, (1991).