日本語複文の意味的構造の解析

1999年2月19日

鳥取大学工学部知能情報工学科
竹内　佐和

目 次

概要

日本語複文は、英語では、関係節、同格節、前置詞句などさまざまな表現に訳されることが多く、これを正しく訳し分けることは、日英機械翻訳の重要な課題の一つとなっていた。日本語複文の意味的構造を英語の構造と対比して分類できれば、翻訳精度の向上につながると期待できる。

そこで本稿では、動詞及び節が底の名詞(英語では先行詞に相当する)を修飾する、連体節を含む文を取り上げ、複分の意味的構造を、内と外の関係、修飾部の陳述度、底の名詞の抽象度、格関係に着目して 16 段に分類した。分類の有効性を、日英機械翻訳機能試験文集 (6,183 文) 中に含まれる複文 432 文を用いて検証した結果、83 %の正解率を得た。

この結果より、内と外の関係、修飾部の陳述度、底の名詞の抽象度、格関係を考慮した複文の分類が、英語と対比した複文の意味的構造解析において有効であることが示された。

1 まえがき

日本語複文は、英語では、関係節、同格節、前置詞句などさまざまな表現に訳されることが多く、これを正しく訳し分けることは、日英機械翻訳の重要な課題の一つとなっていた。日本語複文の意味的構造を英語の構造と対比して分類できれば、翻訳精度の向上につながると期待できる。

従来の複文の分類方法として、山田[1]は、底の名詞(英語では先行詞に相当する)と修飾部の格関係の種類に応じて複文を分類したが、格関係を含まない複文も存在する。そこで、寺村[2]は、底の名詞が連体節内で格関係を有するか否かに着目して、複文を「内の関係」と「外の関係」に分類した。しかし、日英機械翻訳では、これらの分類では、英語表現が一意に決定できない点が問題であった。

そこで本稿では、動詞及び節が底の名詞(英語では先行詞に相当する)を修飾する連体節を含む文を取り上げ、複分の意味的構造を、内と外の関係、修飾部の陳述度、底の名詞の抽象度、格関係に着目して分類を行う。具体的には、まず、日本語の「内」･「外」の関係より英語の「関係節」･「同格節」に、また日本語の修飾部における陳述度「大」･「小」より英語の「節」･「句」に対応させ、複文全体を4つに分類する。そして、「内の関係」については、底の名詞と修飾部の格関係に基づき12に下位分類し、「外の関係」については、底の名詞の抽象度と修飾部の陳述度に基づき4つに下位分類する。そして、この日本語複文の分類における有効性を、「日英機械翻訳機能試験文集」[3](6,183文)より抽出した複文432文を用いて検証する。

以下、第2章では、日英言語における複文構造の違い、第3章では、内と外の関係および陳述度の定義、第4章では、分類の考え方と方法、第5章では、結果と考察、第6章では、今回の結論を述べる。

2　日英言語における複文構造の違い

本研究では、日本語複文の構造を英語の構造と対比して分類するため、日本語、英語の順に、今回対象とする複文の構造について述べる。

2.1　日本語の構造

図1に示すように、修飾部(動詞及び節)が底の名詞(英語の先行詞に相当する)を修飾する、連体節を含む日本語複文を本研究の対象とする。

なお、底の名詞が以下に示す形式名詞については、今回対象外とする。

1. 「とき」、「ため」
   前の節に対しては連体修飾される名詞として働くが、後の節に対しては副詞的な修飾として働くという二面性をもっている。「～とき、」「～ため、」と、接続助詞として考えられるため、今回は取り扱わない。
2. 「こと」、「はず」、「もの」、「の」
   上と同様に考えて、「～ことがある」「～はずだ」と、助動詞化として考えられるため、今回は取り扱わない。

2.2　英語の構造

「日英機械翻訳機能試験文集」[3](6,183文)中に含まれる複文432文を対象に、対応する英語表現の分類を行った。その結果、図2に示すように、関係節33％：同格節8％：準動詞(分詞、不定詞、動名詞)34％：前置詞句18％：その他7％となった。

これより、日本語複文と英語表現の違いについて、以下のことが分かる。

(1) 日本語の連体節は、 英語では必ずしも節には訳されず、準動詞、前置詞句などに訳される場合が半分以上を占める。

(2) 英語で節に訳される場合は、「関係節」、「同格節」のいずれかの構造となる。

3　内と外の関係および陳述度の定義

3.1　内と外の関係について

2章より、英語で節に訳される場合は、「関係節」、「同格節」のいずれかの構造となることがわかった。そこで、複文の分類において、これらの英語表現に日本語複文の意味的構造が対応するように、図3に示す、寺村[2)]の「内と外の関係」法を用いる。

つまり、底の名詞と修飾部の間に格関係を有するものを「内の関係」と呼び、これは、英語の「関係節」に対応する。そして、格関係を含まないものを「外の関係」と呼び、これは、英語の「同格節」に対応する。

3.2 陳述度について

ここでは、日本語複文の分類において、英語の「節」構造と「句」構造への対応に用いる「陳述度」について述べる。

「陳述度」:「文としてのまとまり、完結性の度合。文であるか、ないとすれば文にどれほど近いかを表すもの。」

英語では、主語名詞と定動詞の結び付きが「文」の単位であるという明確な形式的・客観的基準があり、それが他の一部を成すときは「節」となり、不定詞や分詞による「句」の構造とははっきり区別する手立てがある。しかし日本語では、動詞、形容詞、存在詞が言い切りの形で終わればそれで文に成り得るという最低の必要条件はあるが、文と単なる概念を表すだけの動詞についての、形式的基準は求め難い。例えば、「彼がいつも電車の中で読む本は、…」という下線部と、「何か読む本は、…」の下線部とでは、その"文らしさ"において差があることが感じられる。後者は、テンスも場面性も無い形で、英語で言えば不定詞で表すような内容となる。よって、"文らしさ"を表す「陳述度」の概念が、英語の「節」と「句」に対応させるために有効であると考えられる。

本研究では、具体的に、図4に示す陳述度の設定を用いる。これは、寺村[2]の5段階の設定に、陳述度1を現在形、過去形でaとbに分け、また、陳述度2に「~できる」、「~ない」を付加し、英語の「節」と「句」に分類する上での、精度向上を図った。

4　分類の考え方と方法

4.1　分類の考え方

2 章の翻訳例を分析した結果によれば、複文に対応する英語の表現構造は、日本語側から見ると、「底の名詞」と「それを修飾する述部」の性質に強く依存している。具体的には、「内と外の関係」、「底の名詞を修飾する述部の意味的完結性 (陳述度)」及び、「底の名詞の抽象度」の 3 種類の情報が重要な決定要因となっていると推定される。

そこで、以下では、これらの 3 点と「格関係」に着目して、複文の分類を行う。分類の考え方を、図 5 に示す。

(1)　内と外の関係 と 修飾部の陳述度による分類

まず、日本語複文は、「内の関係」・「外の関係」の視点から英語の「関係節」・「同格節」に、また、修飾部の陳述度「大」・「小」の視点から英語の「節」・「句」に対応させることができる。そこで、内と外の関係を縦軸に、修飾部の陳述度を横軸として、複文全体を 4 つに分類する。

(2)　格関係による分類

「内の関係」については、従来の山田[1]の分類方法のように、底の名詞と修飾

部の格関係に基づき、英語の前置詞に対応させ、12 に下位分類する。

(3) 底の名詞の抽象度による分類

「外の関係」について、図 3 の例文「彼が犯人だという噂」と「魚の焼ける匂い」のように、両者の英語表現「同格節」・「前置詞句」の違いが、底の名詞の抽象度および修飾部の陳述度に依存していると言える。そこで、底の名詞の抽象度と修飾部の陳述度に基づき、4 つに下位分類する。

4.2 日本語複文の分類

4.1 の考え方に基づいた、提案する日本語複文の分類を表 1 に示す。標本データとして、「日英機械翻訳機能試験文集」[3)] (6, 183 文) より抽出した複文 432 文を用いた。

(内容については、付録の具体例を参照)

5　結果と考察

5.1　分類の有効性の検証結果

本手法で提案した日本語複文の分類における有効性を、「日英機械翻訳機能試験文集」[3](6,183 文) より抽出した複文 432 文を用いて検証した。

その結果、表 2 に示すように、「内の関係」300 文に対して 89%、「外の関係」127 文に対して 81%の正解率が得られ、複文全体で 83 %の正解率を得た。

なお、機能試験文の日本語複文における英訳が、提案する分類の英語表言と一致した文を正解とする。

5.2　考察

表 1 に示す日本語複文の分類で、英語表現対応に失敗した例を調べた結果、おおまかに以下に示す 4 つに分類できた。

(1) 英語の単語に対応する日本語複文

表 3 に示す、例「私は 見知らぬ 人 から声を掛けられた。→ A stranger called out to me.」のように、日本語の文「見知らぬ人」が、英語では「A stranger」と単語で表現される日本語複文がこれにあたる。このような日本語複文については、データ登録により英語表現を決定できる。

(2) 英語構造「節」と「句」に分類することが困難である日本語複文

表 4 に示す、「本章は 背景とこれに続く 問題点 から構成される。→ This chapter consists of the background and the problems that follow.」などがこれにあたる。この場合、修飾部の陳述度が「～続く」→ 1.a(1) で、英語表現「句」に訳されると提案したのだが、機能試験文では「節」で訳されている。対象データの大半は、陳述度 1.a(1) では英語の「節」に訳されると、分類に成功したのだが、このように失敗例が挙げられたことから、陳述度のみで、英語の「節」と「句」を決定するには、精度上限界があると思われる。

(3) 英訳において文の組み立てが変わる日本語複文

表5に示す、例「一月に売れる電話器は10セットです。→ 10 sets of telephone equipment will be available for sale in January.」がこれにあたる。この場合、英語では「電話器の10セットが一月に売れるであろう。」と単文で表現されており、この英文構造を考慮して英語表現を決定することは、提案した分類のみでは不可能である。

(4) 英訳において話し手の伝達意志が必要となる日本語複文

これは、機能試験文では、次に示す1件しか出現しなかったのだが、翻訳において、精度を下げる重要な要因となると思われたので取り上げた。「これが、私たちがよくその木陰にすわった木です。→ This is the tree in the shade **of which** we would often sit. (→ This is the tree **where** we used often to sit in the shade.)」この場合、話し手の伝達意志により両者の英語表現は使い分けされることになる。前者では、「私たちがよくその木陰にすわった」と「木」が所有関係を示しているのに対し、後者は場所の「デ」格を示している。話し手がどちらの意志を伝えようとしているのか、1文からその情報を取り出すのは困難であり、前後の文との相関関係からの推測など、情報を増やすことが必要と思われる。

6　あとがき

本研究では、日本語複分の意味的構造を英語の構造と対比して、内と外の関係、修飾部の陳述度、底の名詞の抽象度、格関係に着目して 16 段に分類した。この日本語複文の分類における有効性を、「日英機械翻訳機能試験文集」[3](6,183 文) より抽出した複文 432 文を用いて検証した結果、83 %の正解率を得ることができた。この結果より、内と外の関係、修飾部の陳述度、底の名詞の抽象度、格関係を考慮した複文の分類が、英語と対比した複文の意味的構造解析において有効であることが示された。

英語表現対応の失敗例のなかには、話し手の伝達意志が必要となるものなどがあり、これらの扱いが検討課題の一つである。

参考文献

[1] 山田 孝雄：日本文法論 , 宝文館 (1898).

[2] 寺村 秀夫：日本語のシンタクスと意味　~　, くろしお出版 (1982 ~ 1991).

[3] 池原, 白井, 小倉 : ” 言語表現体系の違いに着目した日英機械翻訳機能試験項目の構成” , 人工知能学会誌 Vol.9,No.4,pp.569 ~ 579(1994).