

英日機械翻訳システムにおける 挿入句の処理

武田 紀子

成蹊大学工学部

英日機械翻訳システムにおいて、英文中の任意の位置に、コンマやかっこで区切られて挿入することのできる挿入句を解析し、適切な訳を効率よく求められるようにするには、挿入句を取り出し、その内容を解析し、それが係るものを決定し、訳を生成しなければならない。その処理をシステムに実現するにあたって、いくつかの挿入句を調べた結果、文献により挿入句の使用にある傾向があり、また、挿入句を区切るコンマ、かっこによっても、挿入句の係り受けの関係が決まる場合が多いということが分かった。そこで、一般の文法規則とは独立した簡単な挿入句の係り受けに関する規則を定義し、翻訳対象に合った規則により、的確な挿入句に関する処理ができるようにした。

Processing of Parenthesis in an English-Japanese Machine Translation System

Noriko Takeda

Department of Information Science, Faculty of Engineering
Seikei University

In order for the English-Japanese machine translation system to process parenthesis enclosing with comma or parentheses, it is necessary to find parenthesis in the sentence, analyze it and translate it into Japanese. After studying sentences that include parenthesis, it is found that there is a certain tendency for usage of parenthesis that depends on a literature. So a method to write rules for analyzing dependency on parenthesis is proposed. These rules are easy to write and implement in the system. Consequently, it is possible to offer rules to suit the sentence to be translated. Improvement on correctness and efficiency on analysis of parenthesis is achieved by implementing these rules.

1. はじめに

英文の文章の中に、そこに書かれている名詞や、ある記述に対するより詳しい説明、注釈、言い換え等を文章の流れとは無関係に挿入したい場合がある。これらの説明、注釈等は、コンマ、かっこ等で区切られ、文章中の任意の句の切れ目に、独立して挿入することができる。このような挿入された句全般を以降、挿入句と呼ぶ。挿入句の適切な使用、翻訳は、英文や英日の翻訳に習熟している人にとっても難しい問題である。

英文を日本語に翻訳するにあたって、挿入句を正しく訳文の中に入れるには、以下のようなことをしなければならない。

- ・文章の中から、挿入句を見つける。これは、挿入句が、かっこで囲まれている場合は、容易になれるが、コンマで囲まれている場合は、コンマの用法として、

- ・2つ以上続いた文の区切りを示す
- ・2つ以上の並列した句の区切りを示す
- ・関係代名詞句の継続的用法を示す

等、いくつかの場合が考えられるため、コンマで囲まれている部分が、文章の中でどのような役割をしているかをまず判断しなければならない。

- ・挿入句を生成している句の解析をする。ここでも、コンマで囲まれた挿入句の場合は、挿入句の中にコンマを含む並列関係をもつ句が含まれることもあるので、挿入句の範囲は、一意的には決められない場合がある。

- ・挿入句が、挿入されている理由を見つけ、それが係るもの、説明しているものの範囲を特定する。
- ・挿入句の訳を求め、文章全体の訳の中で、その訳が入る位置を決める。

このような挿入句の処理を英日機械翻訳システムに実現するための問題点としては、どこにでも現れることができ、その構文は多様である挿入句の取り出しをいかに効率良く行うか、挿入句が係るもの、説明しているものの範囲をいかに決定し、適当な訳をだすか、等があげられる。いくつかの科学技術系の文献の挿入句の使われ方を調べたところ、自由に書けるという挿入句の特徴から、筆者、文献により、使われ方にある傾向があることが分かった。そこで、挿入句の解析は、まず、他の文法的範疇の解析と同様に挿入句を取り出し、その内容を解析した後は、他の一般的な構文規則とは独立した係り受けに関する規則に従って、文全体における挿入句の役割を決定することとした。この規則は、他の構文規則とは独立しているので、翻訳対象によって簡単に入れ替えることができる。これにより、より正確な訳が、より効率よく求められるようになった。

本稿では、まず、2. 挿入句の解析、3. 挿入句の係り受けと、役割の決定について述べ、4. で、挿入句に関する規則の記述法、実現の仕方を述べ、

最後に結果について考察する。

2. 挿入句の解析

英文では、文の先頭、中間、最後のどこにでも、文の切れ目となるあらゆる部分に語、句、節を挿入することができる。これらを、挿入句と呼ぶ。挿入句は、コンマ、かっこ、ダッシュ等で区切られている場合もあるし、文の中に埋め込まれている場合もある。本稿では、よく使われるコンマ、かっこで囲まれた挿入句の処理について述べる。

挿入句の解析は、図1. のように行われる。

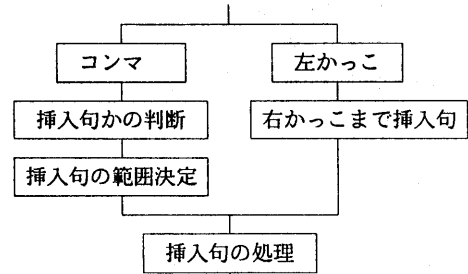


図1. 挿入句の解析

コンマの場合の挿入句であるかどうかの判断は、次のように、キーとなる語の有無を調べる、等によりすることができる。

- ・コンマの直後に、関係代名詞がある場合は、関係代名詞句の継続的用法を生成するとみなされる。
- ・コンマの直後に、接続詞がある場合は、文、句の並列を生成するとみなされる。
- ・文の先頭からコンマまでの語によって文を生成することができず、前置詞句、副詞句等の句を生成している場合は、これらは、文全体に係る挿入句と判断される。
- ・コンマの後に続く語が、" that is, "、" such as " のような語の場合は、挿入句を生成するとみなされる。
- ・文のコンマ以降の部分に、接続詞が含まれていない場合は、挿入句を生成するとみなされる。接続詞が含まれている場合は、まず、コンマで区切られた語によって、並列関係が成立するかの判断がなされ、並列関係が成立しない場合のみ、挿入句を生成するとみなされる。

【例1】

We will frequently have expert systems in mind, in this and next two chapters.

コンマと、接続詞で囲まれた下線部は、その前後の語や句とは並列関係を生成しないので、コンマ以降は挿入句とみなされる。

Along the way the major categories of s-expressions, called lists and atoms, are discussed.

挿入句の範疇	区切り	係るもの	役割	数
名詞句	かっこ	直前の名詞句	省略形	50
名詞句	かっこ	直前の名詞句	同格	35
前置詞句	かっこ	直前の名詞句	修飾	6
前置詞句	コンマ	動詞	修飾	5
過去分詞句	かっこ	直前の名詞句	修飾	4

表1. 文献1における挿入句の使われ方(多い方から、5番目まで)

挿入句の範疇	区切り	係るもの	役割	数
e. g.、i. e. 名詞句	かっこ	直前の名詞句	例示	5
前置詞句	コンマ	動詞	修飾	5
動名詞句	コンマ	動詞	修飾	3
前置詞句	コンマ	直前の名詞句	修飾	3
名詞句	コンマ	直前の名詞句	同格	3

表2. 文献2における挿入句の使われ方(多い方から、5番目まで)

下線の部分は、それぞれ名詞句を生成するので、構文的には並列関係を生成できるが、並列関係の処理から、“lists and atoms”の方が、接続関係が強い、“, called”で、始まる部分は、挿入句に成りやすいとの判断から、コンマ以降は、挿入句とみなされる。

また、コンマではじまる挿入句の終わりの判断は、原則として、次のコンマ、文の終わりでなされるが、挿入句の中に、2つ以上の並列関係や、関係代名詞句が含まれている等の理由により、次のコンマが、挿入句の終わりとならない場合がある。この判断は、挿入句を生成している句の解析の中でなされる。

[例2]

Our aim is to show the essential elements of the language in real programs, but without getting bogged down in details, formal rules and exceptions.

挿入句は、“but without ...”から始まるが、detailsの後のコンマは、下線部の名詞句を解析しているときに、並列関係を生成するものと判断されるので、挿入句は、文の終わりまでとなる。

挿入句が取り出されると、次にそれを生成している部分の解析がなされる。挿入句を生成するものは、名詞句、副詞句、前置詞句等の句が多いが、文や節から成る場合もある。この解析は、文の流れの解析とは独立してなされるが、一般の構文的範疇の解析と同様になされる。

3. 挿入句の係り受けと、役割の決定

挿入句は、ある記述に対する注釈や、より詳しい説明のために書かれる。従って、挿入句の解析を行い、正確な訳を出すためには、挿入句が説明しているものの範囲を決定し、それが使用されている理由

を調べなければならない。文頭にくる挿入句は、文が記述している内容に対する時間、場所、理由等を示す補助的説明であるので、解析は、容易になされる。ここで、いくつかの文献における文の間や文末におかれた挿入句の使われ方を表1.、表2. に示す。

文献1の場合、そこに書かれている146の挿入句を含む文のうち、122文は、かっこで囲まれた挿入句で、そのほとんどが名詞句に係るもので、前の名詞句の省略形を示したり、前の名詞句を別の言い方で表現するために使用されている。

[例3]

The Energy Science Network (ESnet) is
The issues of DSA (Directory Service Agent)
deployment

The ESnet DSAs must operate as Level-0
(first level) DSAs.

これに対し、文献2に含まれる挿入句は、コンマで囲まれるものが多い。(32文中24文)また、文献2のコンマで囲まれた前置詞句からなる挿入句で、名詞句にかかるものは、“such as 名詞句”の形をしたものである。

また、挿入句の係り受けの傾向は、挿入句を生成している部分の範疇、挿入句を囲む記号(コンマ、かっこ)によってもみられる。表3. に前置詞句からなる挿入句の係り受けについて、表4. に、動名詞句からなる挿入句の係り受けについて示す。

ここで、文に係るというものは、2つ以上の文がコンマにより区切られており、その2番目以降の文の先頭におかれた挿入句である場合を含む。また、コンマで区切られた挿入句で、名詞句に係る場合の前置詞句は、“such as 名詞句”という形をしたものである。従って、コンマで囲まれた挿入句は、

区切り	係るもの	数
コンマ	動詞	24
コンマ	文	8
コンマ	名詞	8
かっこ	動詞	8
かっこ	名詞	6

表3. 前置詞句からなる挿入句の係り受け

区切り	係るもの	数
コンマ	動詞	9
コンマ	名詞	3
かっこ	名詞	6

表4. 動名詞句からなる挿入句の係り受け

“主語、挿入句、動詞句”という形をしたものも含めて、動詞に係る場合が非常に多いといえる。いい換えると、前置詞句の係り受けの曖昧性を除くために、前置詞句の前にコンマを付けたともいえる。

コンマで囲まれた動名詞句は、動詞に係る場合が多いが、名詞に係る場合もある。しかし、動名詞句からなる挿入句を含む文のある文献をみると、1つの文献の中に、かっこで囲まれた挿入句を含む文が4文、コンマで囲まれ、名詞句に係る挿入句を含む文、2文が存在していた。従って、この文献の中に現れる動名詞句からなる挿入句は、名詞句に係り、その名詞句のより詳しい説明を与えている場合が多いといえる。

次に、過去分詞句からなる挿入句の係り受けについて、表5. に示す。

区切り	係るもの	数
コンマ	動詞	4
コンマ	名詞	8
かっこ	名詞	4

表5. 過去分詞句からなる挿入句の係り受け

過去分詞句からなる挿入句が名詞句に係る場合は、“called 名詞句”等、同格を表す場合が、多かった。

このほか、挿入句を生成するものとしては、副詞句、to-不定詞句、節や文等があげられる。これらは、動詞に係ったり、文が記述していることに対する注釈を与えたり、それを補完したりするために使用される場合が多い。

また、挿入句を示すための区切り記号による違い

であるが、コンマで囲まれた挿入句は、それが挿入句であることを判断するのが難しいということもあり、使われ方の多様度は多くはない。一方、かっこで囲まれた挿入句の場合は、挿入句の部分がすぐ分かり、容易にその部分を無視して全体の意味がとれるので、様々な使われ方をしている。

[例4]

There are efforts taking place to provide international X400 service over the (international) internet.

There could be two (or more) ESnet trademarks.

しかし、かっこで囲まれた挿入句の場合、それが係るものは、直前、または、直後の語、句である場合が多い。

4. 挿入句の処理の実現

挿入句の処理の実現にあたって、挿入句の取り出し、挿入句を生成している句、節の解析については、[2.] で述べた。次に、自由にどこにでも挿入でき、様々な使われ方をしている挿入句の文全体における役割を決定し、それが係る範囲を決め、訳語、および、その入る位置を決める方法について述べる。しかし、すべての挿入句に対応できる規則を作成するのは、規則が、非常に多く、複雑になるし、また、筆者によって自由に書ける挿入句を限られた規則で表すのは不可能である。そこで、一般的な使われ方に対してのみ、それに関する文法規則を導入し、規則に従って処理を行い、どの規則にも一致しない場合は、前または後の語に係るように処理するという方法が考えられるが、一般的な規則といっても文献、筆者によって異なる場合がある。例えば、動名詞句からなる挿入句であるが、ある文献では、名詞句に係るという使われ方しかしていなかった。

また、訳語の決め方、訳語を置く位置も挿入句の役割によって異なるし、名詞句に係る過去分詞句、動名詞句のように、係り受けが決まっても訳語の与え方が、一様に決まらない場合もある。

[例5]

These include the primary data structure of LISP, called the s-expression.

この文の訳として、一番自然と思われるのは、“これらは、s-式と呼ばれるリスプの基本的なデータ構造を含む。”

であるが、挿入部分を強調するために、

“これらは、s-式と呼ばれるが、リスプの基本的なデータ構造を含む。”

“これらは、リスプの基本的なデータ構造、s-式と呼ばれるが、を含む。”

とすることもできる。

..... these countries (including The United States)

この場合は、

" . . . (USAを含む) これらの国 . . . "

" . . . これらの国 (USAを含むが) . . . "

が、考えられる。これらは、訳者の好みにより選ばれる場合が多い。

そこで、簡単に書き換え可能な挿入句に関する規則の記述法を導入し、文献や、訳者の好みにあった規則を使用することにより、より効率よく、より正しい処理が行えるようにした。これらの規則は、英日機械翻訳システムが書かれているLISPプログラムにより解釈、実行される。

4. 1. 挿入句の係り受けに関する規則

挿入句の係り受けに関する規則は、以下のような内容を持つ5つの要素からなるリストによって表される。

1. 挿入句の前に位置する句の範疇、および、そのキーとなる語の意味カテゴリーまたは語
2. 挿入句を生成しているものの範疇、および、そのキーとなる語の意味カテゴリーまたは語
3. 区切り記号(コンマ(c)、または、かっこ(p))
4. 挿入句に係るもの
5. 訳語の位置

ここで、1、2番目の要素で、キーとなる語の意味カテゴリー、語を指定する必要がない場合は、nil とする。

1番目の要素で指定された名詞句(np)が、主語の場合は、主語(subj)と指定する。

挿入句に係るものの指定は、名詞句(np)、動詞句(vp)、文(S)のいずれかにする。但し、名詞句の場合、かっこで囲まれた挿入句の場合、直前の名詞に係る場合が多い。その場合は、名詞(noun)と指定する。すると、例えば、挿入句の前の名詞句が、名詞句+前置詞句の形をしていても、挿入句は、直前の名詞、つまり、前置詞句を生成している名詞句に係る。但し、いくつかの名詞からなる名詞句の場合は、これ全体が、1つの名詞として考えられるので、名詞と指定されていても全体に係る。

また、訳語の位置は、次のように指定する。

名詞句に係る場合	
名詞句の前の置く	1
名詞句の後に置く	2
動詞に係る場合	
主語の直後に置く	1
動詞の直前に置く	2
文の訳の後に置く	3
文に係る場合	
文の前に置く	1

文の後に置く 2

次に、挿入句の規則の例を示す。

[例6]

((np nil) (prepp such-as) c np 1)

ここで、such-as は、"such as ..." となる場合を表す。"such as" は熟語で、その品詞は前置詞としている。従って、この規則は、名詞句の後のコンマで囲まれた"such as 名詞句" からなる挿入句は、前の名詞句に係り、訳語は、名詞句の訳の前に置かれるということの意味する。

((np nil) (prepp nil) c vp 2)

名詞句の後のコンマで囲まれた前置詞句からなる挿入句は、動詞に係り、その訳語は、動詞の訳の前に置かれる。

((np nil) (gerund nil) p np 2)

名詞句のあとのかっこで囲まれた動名詞句からなる挿入句は、前の名詞句に係り、その訳語は、名詞句の訳の後に置かれる。

4. 2. 係り受けの規則の実現

リストで与えられた挿入句の係り受けに関する個々の規則は、1つのリストにされて英日機械翻訳システムに与えられる。これらの規則が、機械翻訳システムで、どのように利用され挿入句の処理がなされるかを、名詞句の後に挿入句がきた場合を例にとって、図2で述べる。

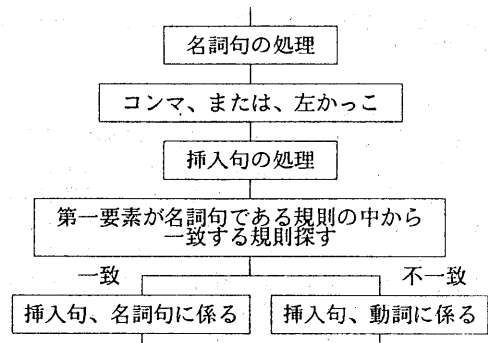


図2. 係り受けの規則による挿入句の処理

挿入句が、名詞句に係ると判断されると、挿入句も含んだ全体が1つの名詞句とされる。

[例7]

((np nil) (gerund nil) p n 2)

という規則で、

.... these countries (including The United States)

を解析すると、

(np (np these countries)

(parenthesis (gerund including The United States)))

となる。

区切り記号	正しく翻訳	誤訳	計
コンマ	7 4	8	8 2
かっこ	2 7	1 1	3 8
計	1 0 1	1 9	1 2 0

表 6. 挿入句を含む文の翻訳結果

また、一致する規則が見つからず、動詞に係ると判断されると、名詞句の処理はそこで終了し、動詞句の処理に戻る。名詞句以外の句の処理中においても同様に、規則のリストの第1要素に登録されている範疇の処理をしているときに、挿入句に出会うと、規則を調べ、係り受けを決定する。また、規則には、登録されていない挿入句の使われ方が見つかったら、それを生成している句、節が、挿入句としてではなくそこに書かれているものとして処理される。

[例 8]

..... two (or more) ESnet trademarks.

このような挿入句の現れ方の規則は、登録されていないので、かっこを含まない文

..... two or more ESnet trademarks.

として、構文解析される。

また、挿入句の規則は、規則のリストに登録されている順序で適用されるので、先に書かれた規則の方が優先度が高くなる。そして、訳語を生成する時に、再び、規則が参照され、訳語の位置が決められ、訳が挿入される。

5. 結果と考察

適当な挿入句の係り受けに関する規則を指定し、挿入句を含む科学技術系の文献に含まれる120文(最大語数 52語、最小語数 8語)の翻訳を行った。規則により、係り受けの決定がなされるので、挿入句、および、その前後の句の意味を考慮して解析する必要がないので、係り受けに関する曖昧性はなくなり、処理もより効率よくなされた。翻訳した結果は、表6. のとおりである。

ここで、正しく翻訳されなかった理由としては、

- ・挿入句の前後の文の解析が正しくなされなかった(挿入句以外の構文解析におけるエラー)
- ・挿入句の訳が正しく求められなかった
- ・係り受けが正しく求められなかった

が、あげられる。特に、かっこで囲まれた挿入句の場合は、かっこという明らかな区切りによって、文の流れとは全く独立して様々な形をした語句が書けるので、挿入句を生成している部分の解析がうまくいかず、挿入句自体を構成しているいくつかの句の係り受け関係等が、正しく求められなかったりして、正しい訳が求められなかった場合が多かった。また、挿入句の訳語生成が、一般の文を構成している句の

訳と同様になされるために、適切な訳ができない場合もあった。

挿入句の使用は、文を書く上で、筆者の個性が出るし、文の読み易さ、理解のし易さにも影響を及ぼす。今回、例として翻訳した科学技術系の文献に現れた挿入句の使用は、いくつかの限られたパターンに含まれるものが多かったが、その他にも多くの使われ方がある。

[例 9]

And of course the mind at, shall we say, eight years old is like a sponge,

前置詞句の前置詞と名詞句の間に、疑問節が挿入されている。

Apart from all the material obstacles (and they are many) there is much that is inimical to the would-be novelist.

and+節(句)が、挿入句として入れられる。

様々な使われ方の可能性を持つ挿入句の翻訳を適切な行うことは、人間にとっても難しいことであるが、より強力な機械翻訳システムを構築するためには、今後、考えなくてはいけないことであろう。

謝辞

日頃、適切なお助言、ご指導を頂いております、成蹊大学名誉教授 和田弘氏に感謝致します。

参考文献

- [1] 久米雅子, 辻井潤一, 長尾真: Muプロジェクトにおける英日翻訳システムの構成と翻訳結果, 情処研報, NL59-6(1987).
- [2] 野上宏康, 杉浦裕美子, 平川秀樹, 天野真家: 英日機械翻訳システムにおける英文の解析方法について, 情処研報, NL61-4(1987).
- [3] 武田紀子: 英文における並列関係の検出, 情報処理学会論文誌, vol. 35, No. 11, pp. 2249-2257 (1994).
- [4] 原沢正喜: 現代英語の用法大成-資料・解釈・評価-, 大修館書店, pp. 489-513(1979).
- [5] Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik: A Grammar of Contemporary English, Longman, pp1058-1072(1972).