

## 衛星放送ニュース文の英語解析文法

6Q-1

鎌田雅子\* 江原暉将\*\* 中瀬純夫\*\*\* 松田健生\*\*\*

(漢字情報サービス\* NHK放送技術研究所\*\* カテナ株式会社\*\*\*)

### 1. はじめに

NHKでは、衛星放送(BS)の英語ニュースのテロップ作成に英日機械翻訳システム<sup>1)</sup>を試用している。テロップ作成上に起きる問題を整理し、辞書、文法にフィードバックし<sup>2)</sup>、精度向上を目指している。翻訳の基本ソフトウェアの文法<sup>3)</sup>をニュース文に対応するよう改良したので、その詳細を述べる。

### 2. BSニュース文と解析規則

基の文法の構文規則は、CFGタイプで、数は3000弱である。曖昧さの解消は、各単語および構文規則のウエイトにより行なっている。また、副詞類は、文法では、かかり先を決めず、フローティングという構造を作って、規則適用後に、辞書情報などを参照してかかり先を決定している。構文規則としては、文法現象を網羅的に取り込んでいる<sup>4)</sup>が、省略、挿入、語順の変化が多い、いわば口語的表現のTVニュース文には、対応できていない。また、テロップ原稿の作成現場から、例えば、ing形を含む単語列を名詞句とした解釈を優先しやすい、などのシステムの傾向や、解析がうまくできない構文などの問題点が指摘された。一方、3000という規則数は、文法を管理していく立場からは「多い」数であり、副作用なく文法改良を行なうために、文法を整理しておく必要があった。

### 3. 文法改良

#### 3.1. 構文規則の組み替え

一般に、構文規則には、入力構文の構成要素を名詞句、動詞句、前置詞句などの句レベルとしてまとめあげるタイプ(句レベル)と、句/文、単文/複文/重文など、句と句の関係をまとめあげるタイプ(構文レベル)がある。しかし、基の文法(G1と呼ぶ)は、句レベルと構文レベルの区別が明確でなく、長い語列を忠実に上位のシンボルに書き換えていく形式であった。構文規則を今後追加していくには、見通しの良い規則の形態が不可欠であるので、

< G 1 >	< G 2 >
(1)SE --> SG PD	(1')SE --> SG PD
(2)SE --> OA CN SG PD	(2')SG --> SG DE
(3)SG --> MX VX	(3')SG --> DE CN SG
(4)VX --> vi5 DB (OA)	(4')SG --> MX VX
(5)OA --> DE	(5')DE --> co2 SG
(6)DE --> co2 SG	(6')MX --> prns
(7)MX --> prns	(7')VX --> vi5 DB (AV)
(8)DB --> av2	(8')PD --> prd
(9)PD --> prd	(9')CN --> cma
(10)CN --> cma	

SE:平叙文, SG:単文, OA:フローティング, DE:副詞節, VX:動詞句, MX:主語名詞句, AV:副詞句, prd:{.}, cma:{.}, co2:{when}, vi5:{go,come}, av2:{out,back}, prns:{she,he}

図1 構文規則の例

規則の適用を、句レベルの解析を行なってから、構文レベルの解析を行なうように変更した。以下、改良した文法をG2と呼ぶ(図1)。

(a) *he went out when she came back.*

(b) *when she came back, he went out.*

(a)(b)は、主節、従属節が入れ替わった構文である。G1では、(a)は、(4)の規則で、*went out when she came back*の語列が動詞句(VX)としてまとめ、主語名詞句(MX)と合わせて、単文(SG)とする。一方、(b)の場合は、(4)の規則で、*came back*、*went out*をそれぞれVXとした後、*he went out*、*she came back*をそれぞれSGとしてまとめ、(2)の規則で副詞修飾(OA)をもった文としてまとめる。

G2では、動詞句内で処理する副詞類は、副詞および前置詞句などの副詞句(AV)に限った。構文レベル規則で、(2')(3')のような主節、従属節間の規則を記述した。複文などの一般的な構文規則は、再帰的な規則適用とし、規則数が増えるのを押さえた。G2では、(a)は(2')、(b)は、(3')の規則で解析する。

規則の組み替えにより句レベル、構文レベルでの処理の範囲が明確になり、規則の改良時に、文法適用の動きが予測でき、作業が容易になった。

#### 3.2. 構文規則の改良

テロップ原稿の担当者から解析失敗の構文、また、翻訳結果に対する要求などが寄せられる。文法に関係する問題点で、主なものは、ing形の解釈誤

English Analysis Grammar for Satellite Broadcasting News  
M. Kamata\*, T. Ehara\*\*, S. Nakase\*\*\*, T. Matsuda\*\*\*  
Kanji Information Service\*, NHK Science and Technical  
Research Laboratories\*\*, CATENA Corp.\*\*\*

り、副詞類のかかり先の誤りであった。

### 3. 2. 1. ing形の解釈

ing形は、動詞の進行形および動名詞、また、多品詞語の場合、名詞、形容詞などの解釈がある。*"They are flying planes."*のように、「飛んでいる飛行機」なのか「飛行機を飛ばしている」のか、統語上の多義もあるが、数の一致や限定詞、形容詞の語順など、文法を精密にすれば判断できる場合も多い。

<数の一致>

(c) *she is baking cakes.*

G1では、(c)に対して、「彼女はケーキを焼いている。」と「彼女は焼けているケーキである。」がほぼ同じウエイトであった。G2では、数の一致によってコプラ解釈の規則を分け、違反時の優先度を低くした。

(c)のG2による翻訳結果

33◇彼女はケーキを焼いている。

35◇彼女はケーキを焼くことである。

124◇彼女は焼けているケーキである。

135◇彼女は焼け付くようにケーキである。

144◇彼女は焼いているケーキである。

bakeは、動詞（焼く、焼ける）、名詞（パン焼き）、bakingは、名詞、形容詞（焼け付くような）、副詞（焼け付くように）の辞書情報を持っている。

<前置修飾要素の語順>

G1は、名詞句解釈を優先する傾向にあった。G1の名詞句のまとめ方が、前置修飾要素の語順を考慮せず、名詞とその修飾要素という2項関係だけで名詞句と判断していたためであった。G2では、限定詞、形容詞類の語順により前置修飾句を先にまとめ、名詞句を生成するように変更した。

(d) *they are arresting several boys.*

G1では、(d)は、「彼らはいくつかの男の子を逮捕している。」とともに「彼らは印象的ないくつかの男の子である。」があった。G2では、後者の解釈がない。

(d)のG2による翻訳結果

24◇彼らはいくつかの男の子を逮捕している。

24◇彼らはいくつかの男の子を逮捕することである。

arrestは、名詞（逮捕）、動詞（逮捕する）、arrestingは、形容詞（印象的な）の辞書情報を持っている。

<動名詞vs.進行形>

動名詞と進行形の区別は、*"his favorite is climbing hills."*（「彼の趣味は丘に登ることである。」-動名詞）*"he is climbing hills."*（「彼は丘に登っている。」-進行形）のように、生物/無生物などの主語の素性で決まる。素性指定の記述はできないので、両解釈が上位に現れるように規則のウエイトを調整した。

### 3. 2. 2. 副詞類のかかり先

BSニュース文には、副詞類は、頻繁に現れ、位置も様々である（(e)(f)の下線部）。

(e) *on the surfaces, it is quiet and tranquil, especially when you compare it with 2 months ago.*

(f) *When I came here, with a psychological examination, they determined type A person.*

G1では、フローティング構造を用いてかかり先の曖昧さによる多義を軽くしているが、間違えも多く、また、別解のかかり先がなかった。かかり先誤りを少なくするには、節、句、単語、とレベルの異なるカテゴリーを一括してフローティング(OA)として扱うままでは、精度の向上は得られない。従って、G2では、かかり先を文法処理し、副詞節、to不定詞、前置詞句、副詞句等の構文規則を記述した。副詞節、to不定詞は、構文レベルで、図1の(2')(3')のように記述した。前置詞句、副詞句に関しては、文頭の句は、主語以降がSGとしてまとまってから、文中、および、文末の句は述部をまとめる時点で任意要素として記述した（図1の(7')）。

(e)のG2による翻訳結果

216◇表面に関して、特にあなたが2月前にそれに匹敵するとき、それは静かで穏やかである。

(f)のG2による翻訳結果

86◇わたしがここに来たとき、心理学の試験で彼らはタイプA人を決定した。

101◇わたしがここに心理学の試験で来たとき、彼らはタイプA人を決定した。

## 4. 結果と考察

規則の改良により、翻訳可能な構文の数も増え、副詞類のかかり先も改善された。G2の規則数は、約3700である。規則の追加により、統語的意味的多義が増加したが、翻訳結果に正解が含まれれば担当者の作業も軽減される。しかし、長文になると、正解の候補の数が増大するので、文を分割するなどの前編集が必要である。今後、時に関する副詞句のまとめ上げなどの前処理を充実させて無駄な多義を排除しながら、翻訳精度の一層の向上を目指す。

### 参考文献

- 1) 相沢、浦谷、田中：放送ニュースへの機械翻訳システムの適用、電子情報通信学会、NLC91-20、1991
- 2) 畑田、江原、山本：衛星放送・英日機械翻訳システムの辞書整備、情報処理学会第46回全国大会、1993
- 3) 中瀬：英日機械翻訳システムにおける解析手法について、情報処理学会、NL67-2、1988
- 4) 浦谷、相沢：英語ニュースの機械翻訳、情報処理学会、NL78-18、1990。