

自然な文章生成のための規範

吉村裕美子 平川秀樹 天野真家

(株)東芝 総合研究所

一般にネットワークやトゥリーなどの二次元表現で現される概念構造から、一次元のデータである表層文を出力する過程で留意すべき点は、「文法」の制約内で、概念構造の持つ意味を失わないまま、自然で理解容易な文章を生成することである。本稿では、その重要なポイントとして、語順操作、パンクチュエーション操作、構文選択の三つの事項を取り上げた。

まず、上記三事項が取り上げられた背景である問題について述べる。次に、これら操作の処理にはどういう情報を用いることができるか、およびその方法を詳説する。さらに、機械翻訳への適用を考えるならば、他にどんな付加情報が利用可能かについても考察し、その処理方法を述べる。

Criteria of Natural Sentence Generation

Yumiko YOSHIMURA Hideki HIRAKAWA Shin-ya AMANO

R & D center, TOSHIBA corporation

1, Komukai Toshiba-cho, Sawai-ku, Kawasaki 210 JAPAN

This paper concerns the way of generating natural, comprehensible sentences from conceptual structures, which are usually represented by grammar-free structures; for instances, network, tree and so on. The authers describe how to control word order and punctuation and to determine appropriate syntactic patterns, in order to avoid unnatural, miscomprehensible, or uncomprehensible sentences. The application to machine translation is also mentioned.

1.はじめに

近年、自然言語処理の研究において、ある概念を表わす抽象構造(概念構造)から自然言語の表層文を出力するための文章生成技術に対する関心が高まり始めた。これまで言語解析技術の向上が基本とされ、主として注力が傾けられてきた。その結果、今日では解析技術も一定レベルに達してきている。解析技術の向上に従い、次にこれを土台・前提とする技術に注目が移行するのは極めて自然なことである。例えば、機械翻訳においては、構文解析能力が向上するにつれ、生成すべきデータの種類は増大し、かつ出力文の質を問われるようになってきた。

概念構造は、一般にネットワークやトゥリーなどの二次元表現で表わされる。一方、表層文は一次元のデータであり、満たさなくてはならない「文法」を持っている。この過程で留意すべき点は、「文法」の制約内で、概念構造の持つ意味を失わないまま、自然で理解容易な文章を生成することである。本稿では、その重要なポイントとして、語順操作、バンクチュエーション操作、構文選択の三つの事項を取り上げた。これに対する対処なくしては、文法的だが不自然な文、誤解を招く文、不可解な文の生成を回避できない。

まず、上記三事項が取り上げられた背景である問題について述べる。次に、これら操作の処理にはどういう情報を用いることができるか、およびその方法を詳説する。さらに、機械翻訳への適用を考えるならば、他にどんな付加情報が利用可能かについても考察し、その処理方法を述べる。

2.これまでの文章生成における問題点

2.1.一つの概念に与えられる個々の語句の持つ素性の違い

一般に、ある言語を生成するための手段として、生成の最小単位となる個々の語句に品詞(名詞、動詞、形容詞、副詞、他)を与えている。すなわち、概念構造中の品詞間の関係を基に表層の単語列を決定するのが従来の手法である。しかし、自然な文章となるよう語順を操作するにはいわゆる文法上の品詞分けでは不十分であり、一つの品詞群に属する語句が個々に持っている様々な素性を考慮した生成処理が必要である。

例えば、英語の副詞(句)を考えてみる。特にコンテキストに依存しない状況で、次に挙げるよう副詞の生起位置の違いが見られる。

- (1) First of all, you have to clear your thoughts.
- (2) He always tries to get up at five.

(3) You must leave here promptly.

これらを全て品詞だけの側面から処理すると、(1)-(2)の生成は例えば次のようになり、不自然な響きや別のニュアンスが加わってくる。

- (1)' You have to clear your thoughts first of all.
- (2)' He tries to get up at five always.

また(4)のように、関係節の中でこのような副詞の生成が行われると、出力文に曖昧性が生まれるどころか意味が歪められて伝わる恐れがある。

(4) The lady with whom you have lunch usually arrives here just before noon.

この文章の場合、シンタックス上では、「usually」は修飾先として“have lunch”と“arrive here just before noon”との二つの候補を持ち、曖昧性を残すが、通常主語と動詞句の間に生起しやすいという“usually”的性質を考えると、後者の解釈が第一義として受取られる。

また、一つの概念に当てられる語句群に、好む生起位置が異なるものが複数混在することもある。(5)-(6)は日本語の「度々」にあたる一つの概念に“often”・“on several occasions”が当てられた際の生成例である。このような生起位置操作は概念構造のレベルでは理論上不可能であり、個々の語句の素性をも取り込んだ処理が行なわれなければならない。

(5) He often visits me.

(6) He visits me on several occasions.

以上の例から明らかなように、概念構造にある意味を失うことなく自然な文章を生成するためには、個々の語句が持っている素性を反映させることが不可欠である。

2.2.生成位置による統語関係の曖昧性

英語の文章において、ある動詞句の中に別の動詞句が埋め込まれると、その後に位置する副詞句のシンタックス上の位置づけは一般に曖昧になる。すなわち、直前の動詞句を修飾するものか、あるいは、その句を飛び越えたより上位の動詞句を修飾するものかという曖昧性である。ところが、人間による文章生成は、文脈から明らかでないかぎり、スピーチであれば、発話のスピードを変えたり、間を空けたり、イントネーションを調節したり、語順を変換したりすることにより、この問題を解消するであろ

う。一方、文語となると、間を空ける代りにパンクチュエーションが効果的に用いられたり、語順の操作が意図的に行われる。次に挙げる(7)-(9)は、以上のような工夫なく生成された文章である。

- (7) I do not know whether his son graduated in medicine though he is my best friend.
- (8) Use the data processed at this stage effectively.
- (9) Print out the date processed at this stage using the command "ymk".

これらは、シンタックス上それぞれ “though ---”, “effectively”, “sing ---” の部分が直前の動詞句 “graduated ---”, “processed ---”, “processed ---” を修飾するのか、その上位の動詞句である “do not know ---”, “Use ---”, “Print out ---” を修飾するのか明確でない。(7)-(9)は、後者の修飾先を暗示させるためには次のような語順操作やコンマの挿入が求められる。

- (7)' Though he is my best friend, I did not know whether his son graduated in medicine.
- (7)" I did not know whether his son graduated in medicine, though he is my best friend.
- (8)' Use effectively the data processed at this stage.
- (9)' Using the command "ymk", print out the data processed at this stage.
- (9)" Print out the date processed at this stage, using the command "ymk".

このような操作は、出力言語の基本文法に従っていてもできるものではない。ある修飾句とその被修飾句との間にどんな構造のどの程度複雑な句が生成されるか、また修飾句の構造はどんなものかという観点に立って初めて可能である。概念構造からの生成の場合、部分的な構造は同じでも、それがどういうサブルーチンの元で生成されるかで語順のストラテジーは大きく異なってくる。よって、生成ルール中で任意にサブルーチンが設けられるのと同程度に柔軟なコントロール機構が必要である。

3.自然な文章生成のための語順操作・パンクチュエーション操作

3.1.個々の語句の持つ素性を用いた語順操作

英語の副詞・形容詞は、個々の語句自身が好む生起位置を持っている。単に、「品詞」という枠組の

みからのカテゴリー化では、語順をコントロールしきれず、概念構造にある意味を失った文章や不自然な文章を生成することになりかねない。そこで、個々の語句に異次元の素性を情報として持たせ、これを生成ルールの各所で参照しながら語順を決定していく方式を採用している。

素性の次元としては、「意味的側面」、「生起位置からの側面」、「形態的側面」の三点をとらえている。また、参照処理を容易にし、処理速度を高めるために、これらの素性を一次元化し、語順操作の効果が最大になるように各語句に値の付与を行なっている。

3.1.1.副詞に対する付与

各種副詞類の生起位置については、Quirk他(1985)[1]では次の七つに分類している。

The book must have been placed
↑ ^ ^ ^ ^ ^
on the shelf.
 ^

1: I
2: iM 3: M 4: mM 5: eM
6: iE 7: E

文中位置は四つに細分されているが、最も一般的である M の位置に生起するものは iM にも同様に生起でき、 mM・eM に生起できるものは M にも生起できる(岡田(1985)[4])。このバリエーションは、否定のスコープやフォーカスにより生まれるもので、個々の語句自身に持たせる情報ほどには確立していない。よって本方式では iM・M・mM・eM を文中位置として一つに凝縮させている。文尾位置についても、同様に個々の語句自身が持つ素性によるというよりは、むしろ各副詞句の持つ修飾のスコープに依存すると考えるのが自然である。(修飾句のスコープを扱う処理については、5節で翻訳との関連で一つのアプローチを解説する。)実際、岡田(1985)[4]では、iE・E の区別を与えていない。本方式でもこれらを文尾位置としてまとめている。かつ、文尾位置を副詞句の生起位置としては無標としている。

統いて、上述の三つの側面からどういう値を取り上げているかを簡単に説明する。実際に各語句に値を付与する際には、語順決定の観点から値に優先順序を設け、高いものを付与している。以下の例挙はその順序に従っている。

●通常、文中位置に生起する。(←生起位置)

ex) He usually doesn't watch TV before din-

- ner.
- 通常、文頭位置の生起する。(←生起位置)

ex) First of all, you have to clear your thoughts.
 - 頻度を表わす。(←意味)

ex) He visits me frequently.
 - 時を表わす。(←意味)

文尾位置に好んで生起する。

ex) I received a letter from her yesterday.
 - 場所を表わす。(←意味)

文尾位置に好んで生起する。「時」と共起するときは「時」に先行する。

ex) They are playing baseball over there.
 - 動詞句から形成される。(←形態)

一般に動詞句はその後に位置する副詞句をスコープとして取り込むため、有標化する必要がある。

ex) He read a book to kill time.
 - 前置詞句から形成されるか、副詞に後置修飾句が付属している。(←形態)

形容詞句を修飾する場合、後置修飾するものをこれに当てる。

ex) The window opened by itself.
 - 無標(その他)

文尾位置に生起し、形容詞句を前置修飾する。

3.1.2.形容詞に対する付与

形容詞類に関しては、主に名詞を修飾する場合の語順操作を考慮して細分を行なっている。前置修飾・後置修飾とともにコンテキストによる各修飾句のスコープに語順は左右されるが、特にコンテキストに依存しない前置修飾句同志の間には相互順序が存在する。本方式では、これを「一般性質」・「大きさ」・「新旧」・「形状」・「由来」・「色」・「材料・材質」・「分詞」・「名詞由来語」に分類した。以下、形容詞句に付与した値の例を優先度順に示す。

- 前置詞句などの後置修飾句が付属している。

(←形態)

名詞を後置修飾する。

ex) peculiar to Japan, known to everyone, etc.
- 名詞を後置修飾する。(←生起位置)

ex) alive, afraid, etc.
- 最上級・比較級の形態をとっている。

ex) oldest, elder, etc.
- 一般的性質を表わす。(←意味)

ex) brave, beautiful, etc.
- 大きさを表わす。(←意味)

ex) big, little, short, tall, etc.
- 新旧・年齢を表わす。(←意味)

- ex) young, old, new, etc.
- 形状を表わす。(←意味)

ex) round, square, etc.
- 由来を表わす。(←意味)

ex) American, Japanese, etc.
- 色を表わす。(←意味)

ex) red, black, blue, etc.
- 材料・材質を表わす。(←意味)

ex) gold, lacquer, wooden, etc.
- 名詞由来語である。(←形態)

ex) animal, etc.
- 現在分詞・過去分詞からなる。(←形態)

ex) living, used, etc.
- 無標(その他)

この様な情報を各語句に持たせることにより、出力言語のみの持つ特性を生成過程という概念構造からも独立した中で操作することができる。また、人間によりある概念に対する語句の選択が行なわれても左右されることはない。

3.2.各句の重さ・構造を考慮した語順操作・パンクチュエーション操作

各句の重さ(生成後の語列の長さ)や句の構造を生成過程で考慮することは、文のスタイルとしてのまとまり(自然さ)、ならびに読みやすさを高める上には欠かせない事項である。このための手段として、生成ルール中の条件記述部に、任意のサブトゥリーを構成する単語数参照、任意のサブトゥリーにおけるアーチ存在チェックを行う記述を可能にした。以下、そのルール記述の例を挙げながら、各句の重さ・構造を考慮した語順操作・パンクチュエーション操作について説明する。

3.2.1.各句の重さ・構造を考慮した語順操作

英語では、end-weightの原理により、重い句・構造の複雑な句は文尾の位置が好まれる。また、新情報は重い構造を持つことが多く、end-weightの原理がend-focusの原理と協同して句の外置を促す。(10)-(11)のような目的語と目的格補語の置換、目的語と副詞句の置換もこの原理から導かれる。

- (10) We made possible a real-time on-line translating communication system connected via communication satellite.
- (11) We cannot set aside a whole system of rules devised by Congress itself.

一般に、動詞句を従える関係節・前置詞句などの長

い複雑な修飾句をサブトゥリーに持つ重い句を文尾以外に置くと、修飾のスコープの問題が絡み、その後に位置する句とのシンタックス上の関係を曖昧にしがちである。上の例の場合、通常の語順で生成される文は、曖昧になるどころかもともとの意味を失ってしまうだろう。

一方、日本語では、長く複雑な句は他の句に先行する傾向があり、文頭の位置に生じやすい。これは、主語や他の句と述語の間に構造の複雑な句が置かれると、同様に修飾のスコープが絡んで、複雑な句に先行する句のシンタックス上の係り関係が判りにくくなり、かつ文のバランス面から見てもぎこちなくなるのを回避する目的が大きい。

- (12) 砂漠の真中で1週間も食料なしでどう生きのびたかという体験談を私は彼から聞かされた。

以上のような語順操作を行なうためには、各句を構成する単語数・各句が従える構造を参照するルール記述が必要である。(10)-(11)を生成するためのルールの記述例を(13)-(14)に示す。

- (13) ---,
OBJ{ pw<8 | pw>6のRELを従えない &
pw>6のNPPを従えない &
---},
COMP,
OBJ, ---

- (14) ---,
ADV{ pw<4 | RELを従えない &
NPPを従えない &
---}[OBJを従える],
OBJ,
ADV, ---

上記2ルールのセマンティクスを簡単に説明する。各項目の記述順が生成順を表している。
'pw'は、概念構造中の指定されたノード以下のサブトゥリーを構成する単語数を示すプロパティー名である。一つのノード(概念)に対し複数語からなる表層語句が与えられていても、それが将来代名詞化される要素を備えていれば'1'、生成を省略される要素を備えていれば'0'とカウントする。(13)により、単語数8以上の目的語(OBJ)、あるいは、単語数7以上の関係節(REL)・前置詞句(NPP)他を従える目的語(OBJ)は、目的格補語(COMP)の後に生成される。そして、(14)により、関係節・前置詞句他を従えない副詞句(ADV)は、目的語の前に生成される。

(12)の生成コントロールもこのようないくつかの条件記述を用いることにより容易である。但し、現時点では、一つのノードに対し複雑な合成語が与えられていて

も、その合成語の構造を情報化しておらず、解析も行なっていないので、「pw」は厳密な重さを反映していない。

- (15) ---,

OBJ{ pw>9 |
pw>4のRELを従える |
pw>4のAPPOを従える |
---} + “を”,
SUBJ + “は”,
OBJ + “を”,
ADV,
v, “。”;

3.2.2. 各句の重さ・構造を考慮したパンクチュエーション操作

各句の修飾のスコープを明確にし、読みやすい文章を得るためにもう一つの重要な手段として、パンクチュエーション操作、特にコンマ・句点の挿入がある。(10)のように、副詞句が複雑でなく簡単な構造であれば目的語との置換が可能であるが、副詞句も目的語同様に複雑な句の場合には、目的語の後にコンマを挿入すると読みやすくなる。また、通常2項の名詞句の並列にはコンマは伴なわないが、(17)のように後項のみに後置修飾句がある場合や、(18)のように前項が後置修飾句のサブトゥリーに更に名詞句を含む場合には、コンマを伴なうことにより並列関係の曖昧性を削除できる。

- (16) I hear the news that the income taxes to be lowered, from the man who attended the meeting with me.
(17) my friend, and the man who attended the meeting with me
(18) the result of the running test, and the source programs

一方、(12)も、更に句点を挿入すれば一層読みやすくなるし、主語も複雑な句であれば目的語を前置せずに主語の後に句点を挿入するのが自然であり望ましい。

- (12)' 砂漠の真中で1週間も食料なしでどう生きのびたかという体験談を、私は彼から聞かされた。
(12)'' 眠りから目覚めたばかりの私は、砂漠の真中で1週間も食料なしでどう生きのびたかという体験談を、彼から無理やり聞かされた。

このコンマ・句点挿入操作も、前節で述べたルール

記述を行なうことにより容易に可能である。(16)を生成するためのルールの記述例を(19)に示す。コンマを挿入する部分を除けば(14)と同じものである。

(19) ---,
ADV{ pw<4 | RELを従えない &
NPPを従えない &
---}[OBJを従える],
OBJ,
","[OBJ{RELを従える |
APPOを従える |
NP{pos=v}を従える |---}
を従える &
未生成のADVを従える],
ADV,---

(17)により、関係節・同格節(APPO)・動名詞句(NP{pos=v})他の後置修飾句を持つ目的語があり、かつ、目的語の前に生成できない副詞句がある場合は、目的語の後にコンマが挿入される。

3.2.3.本操作の問題点

本操作にはまだ問題が残されている。第一に、どういう構造を従える時に何単語以上なら重い句であると言えるかが非常に曖昧である。また、同じ1語でも冠詞・前置詞・代名詞のようにストレスの置かれない語と何音節からもなる名詞とでは感じられ方も異なるだろう。これを簡単に数量化するのは難しい。

二番目として、概念構造中のノードの数からでは現実に出力される語句の正確な量を判定することは困難であるという点がある。実際には、「pw」に与えられる値の他に、生成過程で独自に生成し概念構造にない語句がある。この代表的な例は冠詞で、個別言語に依存しない概念構造には現われない。他にも、概念構造中ではプロパティや概念同志を関係づけるアーク名などに凝縮されているものが表層レベルで具体的な単語として現われるものは同様である。また、逆に、概念構造中に存在しても生成過程で故意に未生成としたり代名詞化するノードもある。例えば、並列される動詞句に繰り返される主語の任意的省略、既出語句の代名詞化が挙げられる。これらを意識したルール記述が必要になるが、記述段階ではその正確な測定は無理である。

これらの問題も、大半は条件部の記述を更に詳細化していくことで克服できると考えられるが、単純に単語の数ではなく、各語句の品詞や音節数によっても句の長さ・複雑さを測定する方向に拡張していきたい。

4.サブトゥリーの重さを考慮した構文決定

前章での操作は、概念構造の個々の概念に出力言語の表層語句が割り当てられた後の操作であった。しかし、実際には、一つの概念を表わすのに、構文のレベルから大きく変わる複数の表現方法が存在する場合がある。例えば、日本語で言えば「---すればよい」にあたる概念の英語表現として次の二つを挙げることができる。

(20) Reading this book is satisfactory.

(21) What is needed is to deliver this letter to his office by tomorrow.

この構文選択の違いは、主としてサブトゥリーの重さとバランス感覚によるものであり、自然な文を生成するためには考慮すべき点である。

これらは、一般ルールでは処理できず、個々の概念ごとに個別に操作されなくてはならない。辞書中の各項目に、サブトゥリーの重さを参照した構文・語彙選択ルールの記述を行なうことで実現される。この種の記述を用いて、次の例で示すような動詞句の表現パターンの選択も、容易にかつ有効的に行なうことができる。

(22) The research planning center provides funds for this national network including some computers in Europe.

(22)'The research planning ceneter provides this national network including some computers in Europe with funds.

'provide'を使った表現としては“provide A for B”と“provide B with A”的2構文があるが、ここでは‘B’にあたる句が重いため、(22)の出力することにより、読みやすくバランスの良い文になる。

しかし、前章で挙げた以上に問題点がある。辞書記述は生成ルールとはステップが異なるため、構文生成過程で行われる独自の調整との呼応がつきにくく、サブトゥリーの重さ判定にさらに不確実性が増すことになる。

5.機械翻訳への応用

5.1.良い訳文と良い文章の違い

根本的には、良い訳文は正確に概念を表現している点で良い文章と言えるはずである。しかし、日本語・英語のように文法体系や文化の異なる言語間の翻訳の場合には、出力言語のみの観点から文章生成を考えると、原文にある意味を曖昧にしてしまうことも起こる。次に、その典型例を一つ挙げる。

(23) The clerk readily agreed to exchange the typewriter, which was obviously damaged.

(24) 店員はタイプライターを取り替えることに快く応じてくれた。それは明らかにこわれていた。

(24)' 店員はタイプライター(それは明らかにこわれていた。)を取り替えることに快く応じてくれた。

(23)は、関係節の継続用法を用いた文章である。これは日本語の文法事項ではない。これを日本語らしく訳すなら、第一の手法として(24)のように関係節部分を次文とすることが挙げられる。日本語を生成するという立場にのみ立てば、これで十分な翻訳である。だが、原文との対応を考える時、2文めの主語にたてた「それ」の使用に問題が生じてくる。代名詞を使用することには、常にその照応行為が伴なう。上の例では、文脈により先行詞は明らかであるが、他での保証はない。つまり、代名詞はそれだけ曖昧性を混入することにつながる。一方、(24)'は、日本語として整っているとは言えない。しかし、「それ」と先行詞が隣接しているため、曖昧性はほとんど生じない。訳文(特に科学技術文献の訳文)としては、後者のほうが適していると言える。

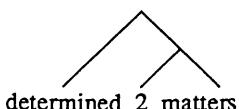
この例からも分かるように、翻訳の結果として良い文章を生成するには、出力言語として自然な文・理解しやすく曖昧でない文を考えるだけでなく、さらに、原文にある意味を最大限保持することにも留意しなくてはならない。次の2節では、そのための一つのアプローチとして、原文中の各句の生起位置情報を訳文生成に取り入れての語順操作について述べる。

5.2. 各修飾句の持つスコープ

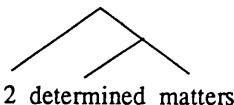
一つの文章の中で、各々の修飾句は、その生起位置により修飾のスコープを異にしている。翻訳において、このスコーピングの状態を最大限生かすことが、望ましい訳文を得る大きな手助けとなる。日本語では、常に修飾句が被修飾句に先行する。一方、英語には前置修飾と後置修飾があり、各々被修飾句の外側に位置する。すなわち、両言語とも被修飾句の格部分から遠ざかるほどスコープが広いと言える。

次の二例は、英語の前置修飾句の位置による意味の違いを示している。

(25)



(26)



(25)は、決定された事項は二つだけであることを暗示している。(26)は、決定された事項が他にもある可能性を消していない。これらに対応する日本語は、例えば、(26)「決定された二つの事項」、(25)「二つの決定事項」が考えられる。この日本語中の生起位置の違いは(24)-(26)の違いに対応している。

上の例からも分かるように、翻訳においては、原文中の各語句の生起位置情報を利用した生成処理が有効であると考えられる。

5.3. 原文中の語順情報を利用した語順操作

概念構造のレベルでは、各サブトゥリーが将来どんな句としてどんな生起位置コントロールを受けるべきか、十分予測されない。そこで、生成ルール中の任意の箇所で、原文中の語順により生成順を操作する記述ができる枠組を考案した。これは、前節で述べたように、一般に被修飾句から離れた修飾句ほど修飾のスコープが広いという原理を利用して、原文における各句のスコープを最大限保持しようというものである。

次に挙げるのは、日英翻訳で、原文の日本語の語順により文頭に生成する副詞句と文尾に生成する副詞句の語順をコントロールするための、ルールの表記モデルである。すなわち、一文中で、前置修飾するものについては原文中の語順の先のものから後のものへ、後置修飾するものについては後のものから先のものへと、順に生成するための記述である。

(27) S = \$PREADV, SUBJ, v, OBJ,
 \$POSTADV;
 \$PREADV = ascend,
 ADV{ 前置修飾条件 },
 ADV{ 前置修飾条件 },
 ADV{ 前置修飾条件 } ;
 \$POSTADV = descend,
 ADV, ADV, ADV ;
 ascend = ;
 descend = ;

'ascend'・'descend'は、形態上、「SUBJ」・「OBJ」などと同じく概念構造中のアーカイブ名の形をとっているが、これらはプログラム上の予約項目である。「ascend」の記述により、以下の処理は、同じアーカイブ名を持つサブトゥリーについては原文中の語

順の先のものから後のものへと生成を進める。逆に‘descend’の記述により、語順の後のものから先のものへと生成処理を行う。この枠組を用いた結果、例えば次の(28)は、各副詞句が前置修飾条件に適合すれば(29)のように生成され、適合しなければ(30)のように生成される。

- (28) 気温の急騰により都心地域では電力の供給が不十分である。
(29) Owing to the sudden rise in temperature, in the midtown area, the supply of electric power is insufficient.
(30) The supply of electric power is insufficient in the midtown area, owing to the sudden rise in temperature.

英日翻訳の場合も同様に、原文中の前置修飾句を日本語として生成する時には‘ascend’を記述して原文の語順通りに出力し、後置修飾句を生成する時には‘descend’を記述して後方の副詞句から順に出力する。

また、原文を解析する際には、必ずしも修飾句と被修飾句との係り関係が一意に決定される訳ではない。(科学技術文献に用いられる文としては適さないが、)(31)においては、「昨日得られたデータ」なのか「昨日行なった再検討」なのか曖昧である。こういった場合、訳文中の語順を操作することにより、原文中の語順が保持する修飾スコープの可能性をそのまま訳文に反映できれば、それはそれで消極的だが一つの解決である。

- (31) 昨日のデータの再検討
(32) reexamination of data yesterday

概念構造上では、一つの動詞句を二つの副詞句が修飾している場合の副詞句同志のように、互いに並列するサブトゥリー間の関係は表されず、差を持たない。この枠組によれば、これら要素に対する語順操作が容易である。

9.おわりに

文章生成の目的は、概念構造にある意味を失わずに、自然な文・理解容易な文を出力することである。概念構造には構造の制約がないが、自然言語には満たすべき「文法」がある。そのため、留意しないと、「文法」は満たしていても、不自然であったり、不可解であったり、誤解を招くような文の生成につながりやすい。本稿では、それを回避する手段として、生成語順の操作およびパンクチュエーションの操作の方式について述べた。その方式において採用したキーは次の三つである。

- 一つの概念に当てられる個々の語句に、異次元の素性を与える。
- 概念構造中の任意のサブトゥリーを構成する単語数、および任意のサブトゥリーの構造を参照する。
- 機械翻訳への応用としては、原文中の語順情報により、各句の修飾のスコープを訳文生成に反映させる。

これらを用いた方式の導入により、シンタックス上の各句相互の関係が明確になり、自然な文・理解容易な文の生成に近づくことができた。

現在、特に機械翻訳のための生成処理に焦点を当てて、生成過程の枠組・ルールの開発を進めている。この際に問題になるのは、5節でも述べたように、良い文章と良い訳文との違いの扱いである。これは、機械翻訳の能力・用いられ方によって次第になくなっていくはずだが、翻訳支援的に用いられることが多い段階では、訳文が原文と対応づけて見やすいこともポイントとなるようである。そのための訳文生成となると、3節で述べたような純粹に生成のための操作と、5節で述べた機械翻訳のための操作とがぶつかることもできる。この相互調整ルールを記述していくと同時に、機械翻訳の訳文に望まれること、純粹に生成処理に望まれることについて、今後も考えていきたい。

また、本文中でも述べているが、サブトゥリーの重さの判定尺度のとして、各語句の品詞や音節数なども取り入れたり、辞書の個々の項目ごとの構文・語彙選択ルールの記述をさらに進めていくつもりである。

参考文献

- [1] Quirk, Randolph et al.: A comprehensive grammar of the English language , Longman, 1985
- [2] Quirk, Randolph & Greenbaum, Sidney : A university grammar of English , Longman, 1973
- [3] Brogan, John A.: Clear technical writing, McGraw-Hill, 1973
- [4] 岡田伸夫：副詞と挿入文，大修館書店，1985