自然言語処理 81−12
（1991. 1. 18）

中間言語からの自然言語生成システム

内田 裕士　朱 美英

㈱富士通研究所

機械翻訳における文生成は，文の木構造表現からのものが多いが，本稿では，国際情報化協力センター(CICC)で進められている近隣諸国間機械翻訳プロジェクトで開発された中間言語（意味ネットワーク表現）から，日本電子化辞書研究所で開発中の電子化辞書の枠組みに基づいた辞書を用い，種々の言語の文を直接的に生成する方法を提案する

A natural language generation system from interlingua

Hiroshi UCHIDA　Meiying ZHU

FUJITSU LABORATORIES LTD.
1015, KAMIKODANAKA NAKAHARA-KU, KAWASAKI 211, JAPAN

Sentence generation on machine translation is mainly from the tree structure representation of sentences. However, in this paper we propose a method to directly generate sentences of various languages using a dictionary based on the frame of electronic dictionaries EDR is developing from the interlingua (semantic network representation) developed by "the machine translation project of among neibouring countries" which are being promoted by CICC.

1. はじめに

ある種の意味表現から，その意味を表す文章を生成するための研究は，機械翻訳における文生成，自然言語による質問応答システムにおける応答文生成，CAI における説明文生成などの目的で盛んに行われてはいるが，解析に比べ，意味のはっきりした情報から出発するということで，力まかせにやれば何とかなるということもあって，研究者の興味を引きにくく，自然言語文の解析に関する研究と比べると低調である．

機械翻訳における文生成は，文の木構造表現からのものが多いが，我々は国際情報化協力センターで進められている近隣諸国間機械翻訳プロジェクトで開発された中間言語[1)2)]（意味ネットワーク表現）から，日本電子化辞書研究所で開発中の電子化辞書の枠組み[3)4)]に基づいた辞書を用い，種々の言語の文を直接的に生成する方法を提案する．

ここで提案する方法は，機械翻訳システムATLAS/U[5)]の文生成部で用いられたもの[6)]を改良したものである．改良の主な外部的要求としては，国際情報化協力センターで進められている近隣諸国間機械翻訳プロジェクトで開発された中間言語を文生成の入力とするということ，および日本の標準的辞書として開発が進められている日本電子化辞書研究所(EDR) の単語辞書および共起辞書と同一の枠組みの単語辞書および共起辞書を利用するということである．

2. 中間言語の特徴

文生成システムの入力である中間言語は，国際情報化協力センターで進められている近隣諸国間機械翻訳プロジェクトで開発されたものであり，文の表す意味を，文の表す概念，その概念に対する話者の視点，概念を表現する話者の意図や概念に対する話者の判断，および文章の構造という観点から表現したものである．

中間言語は，基本的には概念間の二項関係および概念に付加された属性子の集合として表される．この中間言語の特徴としては，

(1)複合概念がひとつの概念としても，また2 つ以上の要素概念から構成された概念としても見ることのできるハイパーネットワークになっているということ，

(2)同一の実体に対してはひとつのノードが与えられ，種々のレベルから参照可能になっている

ことがあげられる．

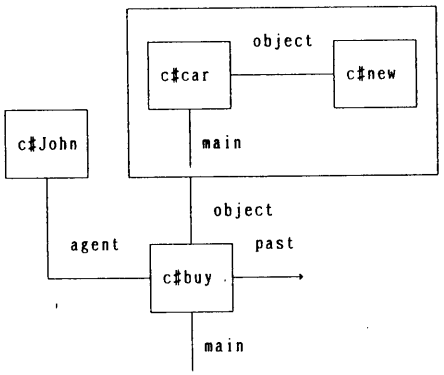

図1 中間言語のネットワーク表現

3. 生成方式

文の意味表現である中間言語から自然言語文を生成する方法としては，意味ネットワークを一次元の単語リストに直接変換する方法を採用した．直接変換するという意味は，文体生成，構文生成，形態素生成が同時並行的に行われるということである．この方法によると，いわゆる疑問変形や受身変形など言語に依存した構造変形や形態変形などを行わなくて済み，言語独立性を高められるという利点がある．

生成システムの入力である中間言語は有方向のハイパーネットワークになっており，ノードが概念等を表し，アークが概念間の関係を示している．

生成システムは，生成窓と出力リストおよび生成規則を解釈実行するルールインタプリタから成る．ルールインタプリタは，与えられた中間言語の各ノードを生成規則に従って生成窓を通して順次に訪れて，結果を出力リストとして持ち帰る．生成システムの構造と使用される規則や辞書類を図2 に示す．

生成窓は，あるノードおよびそのノードに入っているアークと出ているアークを見るためのものである．この窓は生成規則の適用によってノード間を移動していく．出力リストは生成された単語を，その順序で記憶するためのものである．このリストから自然言語文の表層文字列が得られる．

入力された中間言語のノードには概念見出しの他にメッセージバスケットと単語リストが付加される．

文生成の処理は，主ノードから始まり，主ノードが支配する他ノードを巡って主ノードへ戻って終了する．ルールインタプリタは生成規則を解釈し，生成窓を動かし，順次ノードを訪れ，ノードやアークに対す単語を，共起関係や連接関係を参照して正しく選択し，出力リストに

追加していく．生成窓が新しいノードを訪れるとき，もとのノードのすべての状態が保存される．

4.生成システムの動作

生成システムは以下のような手順で中間言語から目標言語の表層文を生成する．

①中間言語を入力し，中間言語を構成するノードのノード名（概念見出し）で示される概念を表現し得る単語を目標言語の単語辞書から検索する．

②中心概念を示すmainというアークが指している概念から生成を始める．

③生成規則にしたがって各ノードを訪問し，目標言語の形態素の候補を決定していく．

④生成された目標言語の形態素の中で連接可能な並びを取り出すことによって目標言語の表層文を生成する．

生成システムは，中間言語を入力してから，まず入力された中間言語に含まれる概念見出しで示される概念を表現し得る全ての単語を目標言語の単語辞書より検索し，検索した単語のリストを各ノード（概念見出し）に付ける．概念見出しで示された概念を表現しえる単語が目標言語に存在しないときは，概念体系を利用し，類義の概念を表す単語があるかどうか，下位の概念を表す単語があるかどうか，上位の概念を表す単語があるかどうかを調べ，対訳語の候補を見付け出す．

5.生成規則

生成規則は，意味ネットワークのノードやアークを横切って，形態素列を得るためのプロダクションルールである．生成規則は生成規則群（生成規則モジュール）としてまとめられている．生成規則群は生成窓から見えるノードとアークに対する生成規則の順序集合である．この順序は生成規則の適用順序を規定しており，結果的に語順を表すことになる．

生成規則の一般形は次の通りである．

```
"/" <condition>  "/" <arc>  "/" <type>  "/"
    <message>  [ ":" <headword> ]  ";"
```

<condition> は，生成規則が適用可能な条件を与える．条件が照合されるのはバスケットに入っているメッセージに対してである．条件が満たされないときは，その生成規則は適用されず，次の生成規則が適用される．

<arc> は生成規則を適用すべきアークを指定する．アークの指定はアーク名およびアークのタイプが指定できる．アークのタイプには，I(インアーク)，O(アウトアーク)，IE(エントリー・インアーク)，OE(エントリー・アウトアーク)，IV(ヴィジット・インアーク)，OV(ヴィジット・アウトアーク）の６種類がある．この欄ではアークの先のノードに関する条件として，概念見出し，単語見出し，文法属性を同時に指定することができる．また，アークのタイプがI,O の場合，２つのノードの候補単語間の構文的役割を示す共起関係子を束縛条件として指定することができる．

アーク指定の形式は，

arc 名（arc タイプ)(共起関係子)

：概念見出し（単語見出し（文法属性））

である．

<type>は生成規則の型を示す．生成規則には次の４つの型がある．

①メッセージの定義を行う型

メッセージの定義を行う型には，ルールが適用されると，<message> に記述されているメッセージが処理中のノードのメッセージバスケットに入るもの(D) と，現在処理中のノードの処理が終了して元のノードに戻るときに，元のノードのメッセージバスケットに送られるもの(M) がある．

②アーク生成を行う型

ノードの周りのアークから始まるサブネットワークに対応する文の生成処理を行う．メッセージ欄のメッセージはアークの先のメッセージバスケットに入る．

③形態素生成を行う型

処理中のノードの概念を表す単語を生成するもの(S)，概念を表す単語以外に必要な単語（例えば，助詞，接続詞，句読点，空白など）を補助出力するもの(G) がある．

補助出力する単語は，文字列，単語見出し，単語グループ名あるいは概念見出しのいずれかの形で指定することができる．また，処理中の単語と補助出力する単語との共起関係による束縛条件も指定することができる．

また，ノードの概念を表す単語を生成するとき，処理中のノードがスコープノードである場合は，この規則が実行されると処理はスコープ内mainアークが指しているノードに移り，<message> 欄のメッセージは main アークが指しているノードのメッセージバスケットに送られることになる．

④生成処理を制御をする型

先読みのために処理中のノードから元のノードへ処理を戻すもの(R) や，指定した生成規則群を呼び出すもの(%)，バックトラックさせるもの(?)，候補単語を変更するもの(!) がある．

<message> では，生成処理をアークの先のノードに移すときに送るメッセージや元来のノードへ戻るときの返送メッセージを指定する．

<headword>では，発生したい形態素をストリング，単語見出し，単語グループ名や概念見出しのいずれかの形で指定する．

生成規則群は各目標言語ごと独立に存在し，GRという名前の特別な生成規則群とその他のいくつかの生成規則群から構成される．

生成システムは，mainが指しているノードから処理を始めるとき，アーク処理のため新しいノードに訪れたとき，あるいはバックトラックや単語変更によってノードの単語が新しい単語に変更されたとき，GR生成規則を起動しする．GR生成規則では，ノードの回りの環境（回りのアークとその先にあるノードなど）および処理中の単語情報（文法属性など）によって，どの生成規則群で生成するかの判断を行い，その生成規則群を指定する．以降，この指定された生成規則に書かれているルールに従って，生成システムは動作し，目標言語文を構成する各形態素の候補を決定し，最終的に各形態素候補より連接可能な並びを連結して表層文を生成する．

6.訳語選択

中間言語から目標言語の文を生成していくときに，どのように訳語を選択するかは，機械翻訳システムにおける文生成の最重要課題である．本生成システムにおいては以下のような方法で，訳語選択を行うこととした．

6.1 概念体系利用による訳語抽出

訳語選択の第１ステップは，まず訳語の候補を見付けだすことから始まる．これは中間言語に含まれる概念見出しで示される概念を表現し得る単語を目標言語の単語辞書から探し出すことによって行われる．概念見出しで示された概念を表現しえる単語が目標言語に存在しないときは，概念体系辞書を利用し，類義の概念を表す単語があるかどうか，下位の概念を表す単語があるかどうか，上位の概念を表す単語があるかどうかが調べられる．類義の単語があった場合，それらが訳語の候補となる．類義の単語がなければ，下位，上位の順に訳語の候補が選ばれる．

6.2 構文的束縛による訳語選択

訳語の候補が見付けだされた後に，生成規則の適用が始まる．ここで中間言語の内容に応じて，その内容を表現するための構文が決定されて行くが，この過程において各ノードに対する構文的役割が決められて行き，それが束縛条件となって訳語が選択されて行く．

6.3 共起関係による訳語選択

同じ概念を表現し，同様の構文的役割を果たせる訳語は一般的に複数個ある．この複数個の訳語の中から適切な訳語を選択するためには，その言語に依存した（固有の）言葉の言い回しに関する情報が必要となってくる．共起（関係）辞書はこのような語用に関する情報を与えるためのものである．これは，ある単語が，どのような関係で，どういう単語と共起するのかという共起関係を定義したものである．この情報は，ある意味内容を自然言語で表現しようとしたときに，他の単語との関係において，ある概念を表す適切な単語を選ぶために使用されるものである．

例えば "c#直す" という概念を表す英語の単語は "modify","correct","update","mend"など複数個存在する．この中から "c#直す" という概念の対象になるものが何であるかということによって適切な訳語を選ぶために用いられる．例えば "c#直す" の対象が "c#エラー" であれば"correct" という単語を選ぶということは "correct" と "error"が対象といういう関係で共起するという情報を用いて行われる．

6.4 連接関係による形態素選択

連接関係は，ある単語がどの単語と隣接し得るかを示す関係である．共起関係が単語と単語の構文的関係であるのに対し，連接関係は単語の並びに関する関係である．したがって単語の並びに関する束縛による訳語選択は連接関係を用いて行われる．

7.省略語（概念）の生成

入力文に何らかの形での省略があったり，また概念を直接的にはもたない単語によってある概念が暗示されているような表現があった場合には，中間言語の表現において，その概念が明示されないことがある．このために，生成システムにおいてそれらの概念を補って，適切な単語を生成する必要がある．

7.1 空概念の生成

入力文に省略があった場合の中間言語は，省略された部分の概念が補われる形で作られることがある．

例えば，日本語「一匹が来る」の場合は，文に含まれる単語のもつ概念間に意味的な関係を与えようとすると，新たに概念を補わなければ表現できないことになる．
概念を補うときは，補うべき概念が明らかな場合と，その概念の持つ属性（上位概念）しか推定できない場合がある．上記の例は，後者のケースである．

中間言語のノードが表す概念は，他の概念との意味的な関係によって，その概念の表す属性が束縛されること

になる．このために，省略された語の概念について属性が推定できるような場合は，わざわざ概念辞書[7]を用いて属性を表す概念を推定し中間言語に表現せずに，，空の概念で表現することがある．

このような，他の概念との関係によって推定できる概念に対しての表層表現を生成すためには，他の概念との概念関係から空概念が何であるかを概念辞書より推定し，この概念に対応した単語を単語辞書より検索して，生成を行う．

7.2 助数詞の生成

助数詞は，物やことを計る単位として用いられるため，助数詞自身が概念を持たないものと考え，中間言語では，物やこと概念は直接数量概念と関係付けられる．

このような中間言語から自然な文を生成するためには，助数詞を発生する必要がある．どの助数詞を使用するかは単語（名詞）によって異なる．このような情報は，言葉の言い回しに関する情報として共起辞書に与えられることになる．

例えば，図3 に示すような共起辞書を用意しておけば，「本」という単語に対して，グループ名 "G#助数詞" を指定すれば「冊」が生成され，「鳥」に対してグループ名 "G#助数詞" を指定すれば「羽」が生成される．

```
(本     ,  @助数詞 ,  冊) = 1;
(紙     ,  @助数詞 ,  枚) = 1;
(鳥     ,  @助数詞 ,  羽) = 1;
(紙     ,  @助数詞 ,  枚) = 1;
 ・・・
(G#助数詞 , @SUPW ,  冊) = 1;
(G#助数詞 , @SUPW ,  枚) = 1;
(G#助数詞 , @SUPW ,  羽) = 1;
(G#助数詞 , @SUPW ,  枚) = 1;
 ・・・
```

図3 助数詞に関する共起辞書例

8.文脈生成[8]

人間にとって自然な文章を生成するためには，中間言語にある概念が示されているからといって必ずしもその概念を表す単語を生成してよいとは限らない．例えば，その概念そのものを表す単語を省略したほうがいい場合や，代名詞を用いて生成した方がよい場合も多い．

このために，ノード間のメッセージ返送機能を用いて文の間でメッセージを転送する機能が用意されている．この機能を用いて文脈を考慮した生成が可能になっている．

9.おわりに

本論文では，意味ネットワーク形式の意味表現である中間言語から自然言語を生成するために必要な機能と情報を示した．ここで提案している文生成システムの前身であるATLAS/U の文生成システムは日本語，英語，ドイツ語，フランス語，中国語等の言語生成を行えるようになっており，生成規則および生成メカニズムの言語独立性は実証されている．本文生成システムは，より表現能力の高いCICCの中間言語からの言語生成を，EDR と同一の枠組みの辞書を用いて行うための一つの方法を示したものである．

参考文献

1)内田裕士，朱美英：多言語翻訳のための中間言語の構成法，自然言語処理研究会資料，72-9(1989.5.19)

2)Uchida,H.,Zhu,M.:Interlingua, Manuscripts of International Symposium on Multilingual Machine Translation'90(1990.11.5)

3)EDR 電子化辞書，TR-17,日本電子化辞書研究所(1990)

4)Uchida,H.:Electronic Dictionary, Proceedings of International AI Symposium 90, pp89-94(1990.11)

5)内田裕士，小部正人，西野文人，増山顕成，松井くにお：日英機械翻訳システムATLAS/U,自然言語処理研究会資料，29-3(1982.1.22)

6)内田裕士：概念構造からの自然言語文生成，自然言語処理技術シンポジウム論文集，pp.57-65(1983.6.16)

7)概念辞書，TR-20，日本電子化辞書研究所(1990)

8)内田裕士，安藤進：文脈情報を用いた自然言語文生成，情報処理学会第29回全国大会講演論文集，4N-7，pp.1235-1236(1984.9.12)

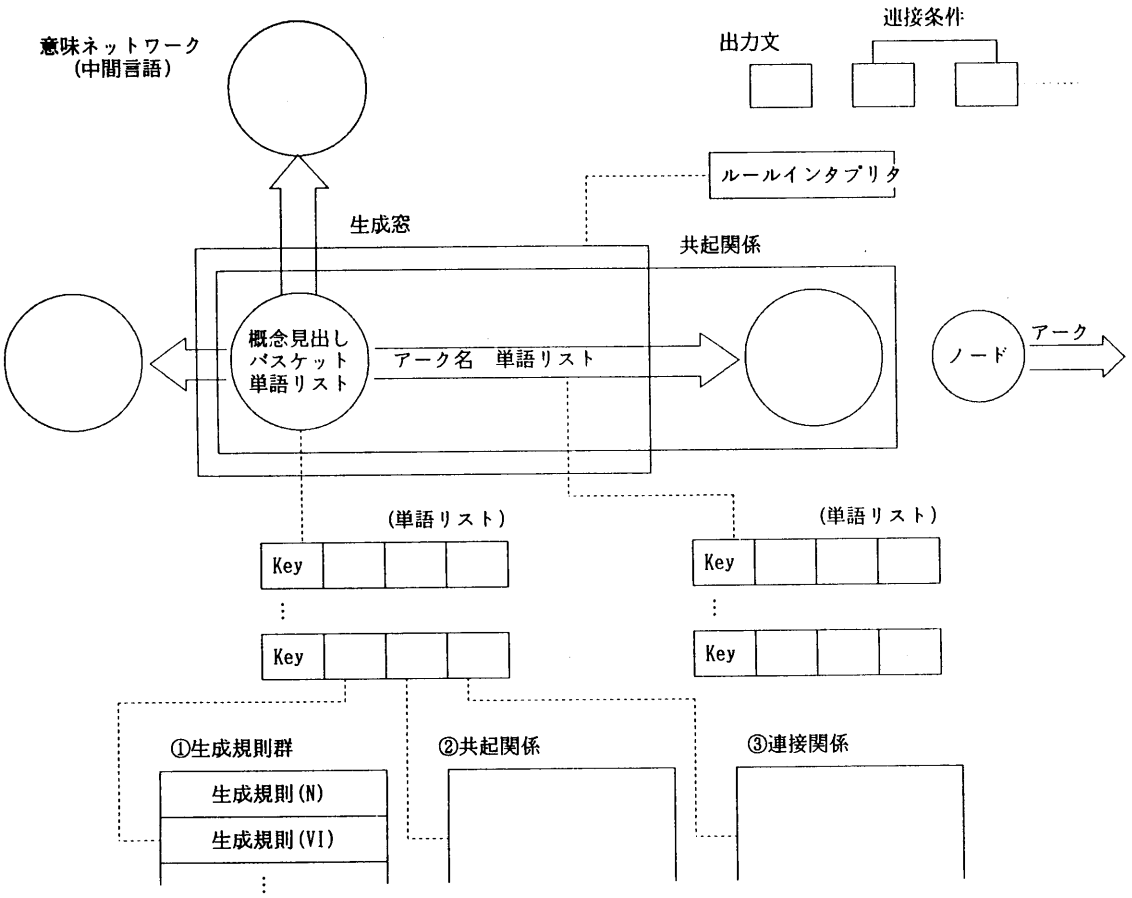

図2 生成システムの構成