自然言語処理 93－2
（1993. 1. 13）

外電ニュースの定型文抽出とその英日機械翻訳

加藤 直人　　相沢 輝昭

NHK放送技術研究所

大量の外電を翻訳するために，機械翻訳の研究を進めている．外電ニュースは多岐にわたっているが，分野を限ると定型的な言い回しが多く使われるものもある．そこで外電のニュースをデータベース化し，英語ニュースの分析を進めている．その第一歩として２～１０語の連続語の頻度を計算した．

本稿では，この連続語を使って，定型文を抽出する方法について述べる．定型文は，これらのみを特別な方法で翻訳する方が精度の向上が期待できる．そこで，定型文を翻訳する方法と，その際に使われる定型文翻訳データを自動的に登録する方法について述べる．定型文翻訳方法を取り入れた，外電経済ニュース英日機械翻訳システムとその評価について報告する．

Extraction and machine translation of sentences with fixed patterns for AP wire service news stories

Naoto KATOH　　Teruaki AIZAWA

NHK Science and Technical Research Laboratories

This paper proposes an extraction method of sentences with fixed patterns and a translation method of the sentences.

We have been developing English-to-Japanese machine translation system for wire service news stories. To improve the system, the stories have been analyzed and 2 to 10 sequential words appearing most frequently were extracted from them.

Our extraction method uses the sequential words. The sentences are better translated by our translation method using features of fixed patterns than conventional machine translation methods. We applied the translation method to English-to-Japanese machine translation for economic news stories and evaluated it.

1．はじめに

毎日大量に入ってくる外電に対応するために，機械翻訳のニーズが高まっている．外電で送られてくるニュースは政治，経済，スポーツ等多岐にわたっており，使われる語や言い回しも様々である．しかし，分野を限ってみると定型的な言い回しが多く使われるものもある．

特に，経済ニュースに関してみると，

1）特殊な文型を持っている．

2）高品質の翻訳を得るためには特別な日本語訳が必要である．

等の特徴がある．これは機械翻訳にとっては，

1）文法ルールが足りない．（しかし，単純に増やすと構文的曖昧性が増える．）

2）訳語選択の困難さ．

という問題に対応する．これらは非常に解決困難な問題であり，実際に翻訳率も低い．（我々の翻訳システムの場合，約２０％である．）しかし，これらの特徴を分析して利用すれば翻訳率の向上が期待できる．

我々はニュース文の特徴をつかむために，ＡＰ電のニュースをデータベース化し，英語ニュースの分析を進めている．その第一歩として２～１０語の連続して出現する語（連続語）の頻度を計算した．

本稿では，このデータの中である頻度以上出現する連続語を使って，定型パターンを含む文（定型文）を抽出する方法について述べる．

このような定型文は，従来の機械翻訳のような解析―変換―生成という過程を経て翻訳するよりも，これらの文のみを特別な方法で翻訳する方が精度の向上が期待できる．そこで，機械翻訳の前段階で定型パターンを含む文を翻訳する方法について述べる．外電の経済ニュースに的を絞り，定型文翻訳方法を取り入れた，外電経済ニュース英日機械翻訳システムとその評価結果について報告する．

2．定型文の抽出

2．1　定型文の抽出方法

我々は以前，あらかじめ数字や曜日などの同一化処理を行なった後，２～１０語連続の単語列に関する頻度データをとることにより定型パターンを抽出した．さらにこれらの定型パターンを連結することにより，１１語連続データ，１２語連続データ，・・・を作り定型文を推定した[1]．この方法では文中に定型パターンを多く含んでいる文であっても，同一化処理にもれた語や１語でも低頻度の語が途中に含まれていると，類似している文であっても定型文として推定できない．図１のような文では，出現頻度が高い語"rose", "fell"を含む1)や2)を抽出することはできるが，出現頻度の低い語"gained", "slipped", "edged up"を含む文は抽出できない．

1) The NYSE's composite index *rose* 0.39 to 196.61.
2) The NYSE's composite index *fell* 0.96 to 183.74.
3) The NYSE's composite index *gained* 0.36 to183.21.
4) The NYSE's composite index *slipped* 0.05 to 163.59.
5) The NYSE's composite index *edged up* 0.33 to 186.51.

図１　定型文の例

そこで，ある頻度以上出現した連続語をデータとし，これらの連続語が１文中に何パーセント含まれているか，すなわち，

含有率＝連続語の語数／総語数

を各文に対して計算した．（ただし，数字の同一化処理は行なった．）この中からあるパーセント以上の文を定型文として抽出した．

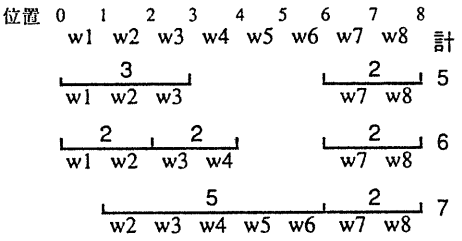
含有率を計算する際に文中のどこで連続する語を区切るかによって，含有率が変わってくる．例えば文として

"w1 w2 w3 w4 w5 w6 w7 w8"　（wiは単語）

を考え，その中に含まれている語の中で，

"w1 w2 w3", "w1 w2", "w2 w3 w4 w5 w6",
"w3 w4", "w7 w8"

が連続語データとして登録されているとする．すると，文中の連続する語の区切り方によって図２



のような３つの場合が生じる．

図２　語の区切り位置

ここで，語wiと語wi+1の間に番号iを付け，この番号を位置と呼ぶ．ただし，文中の最初の語の左側を位置０，最後の語の右側を位置ｎ（ｎは文中の語数）とする．

含有率を計算する際には，一番多くの語数を含む区切り方が適切である．図２の場合では３番目の７語の場合である．そこで次のようにして，一番多くの語数を含む区切り方を選ぶ．各位置間の点数を，その位置間に挟まれる語列が連続語のデータに含まれて入ればその語数と，含まれていなければ０と定義する．ダイナミックプログラミングにより位置０からｎまでの点数の和の最大値を求める．すなわち，位置ｉからｊ（ｉ＜ｊ）までの点数 point(i, j) とし，

$$\text{point}(i, j) = \begin{cases} j-i & \text{if } w_{i+1} \ldots w_j \text{ is in data} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{score}(j) = \max_{0 \le k < j} \left(\text{score}(k) + \text{point}(k, j) \right)$$

で score(n) を求める．得られた値を一文中の連続語の語数とし含有率を計算する．図２の例では図３のような位置間の点数を持つ場合の最大値を求めればよいので，７が連続語の語数となり，

含有率＝７／８＝0.875

となる．

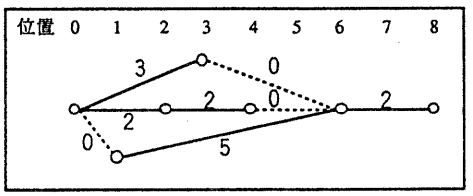


図３　位置間の点数

本方法では連続語を使うことにより，文中に出現頻度の低い語を含んでいても，その語の前後が定型パターンであれば定型文として抽出できる．図１の例では，"The NYSE's composite index"，"数字to数字" が定型パターンであるので3), 4), 5)も定型文として抽出できる．また，１単語の高頻度語を使わないことにより基本的な単語（定冠詞，前置詞，be動詞等）のみが含まれいて，定型パターンを含まない文を排除できる．もちろん，

・何語連続の語を使うか
・どのくらいの頻度以上のものにするか
・含有率を何％にするか

というパラメータを変えることにより，抽出される文を調整することができる．

２．２　定型文の抽出実験

本方法で約２年間分のＡＰ電（総文数約１６０万文）から延べ約２.１万文（数字同一化後の異なり文は８千文）の定型文を抽出した．ただし，パラメータとしては，３～６の連続語を使い，その頻度を１０回以上とし，またディスクの容量の都合で含有率を８０％とした．

抽出した定型文の一部の例を付録に示す．ここで，各文の先頭の数は含有率を表す．実際の結果でもこの例同様，ほとんどが経済ニュースに関するものであった．

３．外電経済ニュースの機械翻訳

２.で抽出した定型文の上位には経済ニュースが多かった．経済ニュースの翻訳は非常に難しい．これは，経済独特の英語表現が含まれ複雑な構文を持つ，日本語訳にも経済固有の訳が要求される，等の理由からである．しかし，経済ニュースは数量表現，曜日などが変化するだけという文も多く，一般のニュースに比べると文の種類ははるかに少ない．したがって，経済ニュースに対しては従来のような機械翻訳は必要なく，単純に訳語の置き換えで翻訳できる場合が多い．

以下では，この特徴を利用した定型文の翻訳方法とその際に使われるサンプル定型文を自動的に登録する方法について述べる．はじめに定型文翻訳に深くかかわる局所解析処理について，簡単に説明する．定型文翻訳では，数量表現等を翻訳する局所解析処理で，文中の変化部分を翻訳する．また，この局所解析処理を使って定型文を翻訳する際に使われる定型文翻訳データを，サンプル定型文から作成する．

３．１　局所解析処理

我々の翻訳システムでは，数量表現，日付・時刻表現，固有名詞等は構文解析に先立って局所的に処理が行なわれ，新しい形態素候補が作られる[2]．例えば，

［例文１］

"In Kuala Lumpur, Malaysian tin closed at 17.76 dollars per kilo, up 5 cents"

の局所解析処理を考える．この文の形態素解析が終了した後，局所解析処理で数量表現

"17.76 dollars per kilo"

"5 cents"

はそれぞれ，

日本語訳	品詞	意味マーカー
「1キロ17.76 ドル」	名詞	数量表現
「5 セント」	名詞	数量表現

と形態素候補ができ，構文解析に送られる．

この処理は構文解析とは別のCHART法[3]によって処理される．解析はＣＦＧルールによって行なわれるが，変換・生成も同時に行ない，ルールが適用されたときに同時に日本語訳を作成する．局所解析処理用辞書と文法ルールを局所解析データと呼ぶ．英語の解析ルールとそれに対応する日本語テンプレートの例の一部を図４に示す．

ここで，"S"は開始記号，"UNTEXP"，"NUMEXP"は非終端記号，"UNIT"，"ABOUT"，"PER"，"NUM"は前終端記号である．また，#n#は解析ルール右辺第n項に対応する日本語訳を表す．例えば，４番目のルールでは#1#は「約」を表す．以下では，このｎをルール位置と呼ぶ．

英語解析ルール	日本語訳
S --> UNTEXP	「#1#」
UNTEXP --> UNTEXP PER UNIT	「1#3##1#」
--> NUMEXP UNIT	「#1##2#」
NUMEXP --> ABOUT NUMEXP	「#1##2#」
--> NUMEXP	「#1#」
UNIT --> "dollar", "kilo", etc.	「ドル...」
ABOUT --> "about"	「約」
PER --> "per", "a"	「」
NUMEXP --> "1", "12", etc.	「1...」

図４　局所解析データ

３．２　定型文の翻訳方法

定型文の翻訳方法は，基本的には局所解析処理を使っている．すなわち，定型文翻訳の処理系は，局所解析処理と同じものを使い，後述する定型文翻訳データを用いて局所解析処理の文法や辞書を拡張し，定型文を翻訳する．数量表現等は前述した局所解析処理で翻訳される．

英語の解析は文ごとにＣＦＧルールで表わした．しかし，英文法に基づくものでなく，単なるパターンマッチングとしている．また，日本語生成は英語の解析と同時に行ない，文単位に対応する日本語訳をあらかじめ与え，局所解析処理で数量表現等変化する部分のみを置き換える．例えば例文１の場合には，数量表現等以外の部分は定型パターンとし，英語の解析ルールおよび日本語テンプレートは図５のようになる．

S --> PAT1 CMA PAT2 UNTEXP CMA UPDW UNTEXP
「クアラルンプールでマレーシアのすずは，#7##6#の#4#でひけた」

PAT1 --> "In Kuala Lumpur"	「」
CMA --> ","	「」
PAT2 --> "Malaysian tin closed at"	「」
UPDW --> "up"	「アップ」
UNTEXP --> "17.76 dollars per kilo"	「1キロ17.76 ドル」
--> "5 cents"	「5 セント」

（UNTEXPは局所解析処理による）

図５　例文１の英語解析

以下では，このような英語解析ルールと日本語テンプレートを定型文翻訳データと呼ぶ．この定型文翻訳データは少数のサンプル定型文から後述するように自動的に作成される．定型文翻訳データは具体的には図５に示すような日本語テンプレートを持つＣＦＧルールと，定型パターンPAT1やPAT2等を含む辞書である．

さて，局所解析処理用辞書に

UPDW --> "down"　「ダウン」

と登録されていると，構文は同じで数値と"down"のみが異なっている，次の例文

［例文２］

"In Kuala Lumpur, Malaysian tin closed at *18.49 dollars per kilo, down 18 cents*"

もまた，

「クアラルンプールでマレーシアのすずは，*18 セント ダウンの1キロ18.49 ドル*でひけた」

と翻訳できる．

３．３　定型文翻訳データの自動作成

本定型文翻訳方法ではサンプル定型文の数だけ英語解析ルールを作る必要がある．また，日本語訳の中で変化する部分を修正した，日本語テンプレートを作成しなければならない．しかし，人手によってこれらを大量に作るのは煩雑である．一方，文法を考慮していないので，このようなルールを自動的に作るのは比較的簡単である．

我々は英文に対して対訳の形で日本語訳を人手で与え，局所解析処理を用いて自動的に定型文翻訳データを作成した．アルゴリズムを次に示す．

［アルゴリズム］

STEP1 文 w1...wn を局所解析処理

STEP2 図６によりコストを決定.

STEP3 ダイナミックプログラミングにより，最も変換する語数が多い場合を選択.

STEP4 局所解析されなかった英語単語列に対して適当な前終端記号を与え，英語解析ルールを作成.

STEP5 対訳中，STEP1で得られた日本語訳は，英語解析ルール位置の番号に置き換え，日本語テンプレートを作成.

```
for i := 0 to n-1 do
  for j := i+1 to n do
    if 位置 i と位置 j の間に非活性エッジがあり，
     その日本語訳が対訳中にあるか
     then
       path(i, j) = 2^(j-i)
```

図６　コストの計算

STEP1では文 w1...wn の局所解析処理を行ない，解析途中に非活性エッジとして，様々な数量表現や辞書に含まれる語列と，その日本語訳が得られる．STEP2では非活性エッジに対応する日本語訳が対訳中にあるかないかを判定し，位置間のコストを計算する．このとき，定型文抽出のときと同じように様々な区切り方があるが，含まれる語数が最も多いように，STEP3でダイナミックプログラミングにより選ぶ．STEP3で処理されなかった語列に対して，STEP4で適当な前終端記号を与え，英語解析ルールを作成する．STEP5では対訳中，STEP1で得られた日本語訳は置き換え，日本語テンプレートを作る．その際には英語解析ルール位置の番号に変換する．

例文１では次のように英語解析用ルールと日本語テンプレートが自動作成される．例文１に対して日本語訳，

「クアラルンプールでマレーシアのすずは，５セントアップの1キロ17.76 ドルでひけた」

を人手で与え，入力とする．STEP1で局所解析処理が終了すると，解析の途中結果の中に非活性エッジとして図７が得られる．STEP2により区切り方および経路間コストは，

","	2
"17.76 dollars per kilo"	16
","	2
"up"	2
" 5 cents"	4

等と決定される．最大値は

２＋１６＋２＋２＋４＝２６

と求められる．STEP3で処理されなかった語列には，STEP4でそれぞれに適当な前終端記号を自動的に作り，

PAT1 --> "In Kuala Lumpur"

PAT2 --> " Malaysian tin closed at"

とする．これらは実際には辞書に登録される．したがって，前終端記号や非終端記号を順番に並べて，英語解析ルールは，

S --> PAT1 CMA PAT2 UNTEXP CMA UPDW UNTEXP

と得られる．このとき各記号のルール位置が決まる．STEP5では対訳中にある日本語訳をこのルール位置に変換し，

「1キロ17.76 ドル」	#4#
「アップ」	#6#
「5 セント」	#7#

となる．よって，このルールに対応する日本語テンプレートは

「クアラルンプールでマレーシアのすずは，#7##6#の#4#でひけた」

と得られる．

このようにして得られた定型文翻訳データと，従来からの局所解析データを合わせてできる，辞書とルールを使って定型文を翻訳する．

定型文翻訳データを作成する際に，単純に数字のみを置き換えずに，数量表現を変数にしたのは次の２つの理由による．

１）原文とは異なった数量表現でも翻訳できるようにする．例えば，

"In *Tokyo*, the dollar fell *0.77 yen* to a closing 143.70 yen."

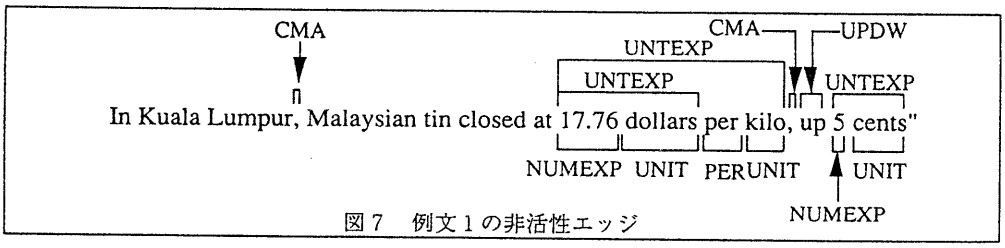


図７　例文１の非活性エッジ

が登録されていれば

"In *London*, the British pound fell *0.005 dollars* to a closing 1.6935 dollars."

も翻訳できる．

2）数字のみにすると，同じ数字が1つの英文中に2度出現した場合，どの数字がどの日本語訳に対応するか曖昧になる．例えば，

"... Tuesday, off 0.74 points or 0.74 percent from Monday's ..."

「...0.74 ポイント（0.74パーセント）...」

では，数字のみにすると対訳中の日本語訳を捜す際に「0.74 ポイント」の 0.74 なのか，「0.74パーセント」の 0.74 なのか曖昧になる．しかし，数量表現とすることにより，対訳中「0.74 ポイント」や「0.74パーセント」を置き換えることができる．

図8に定型文データ自動作成システムの概要を表す図を示す．

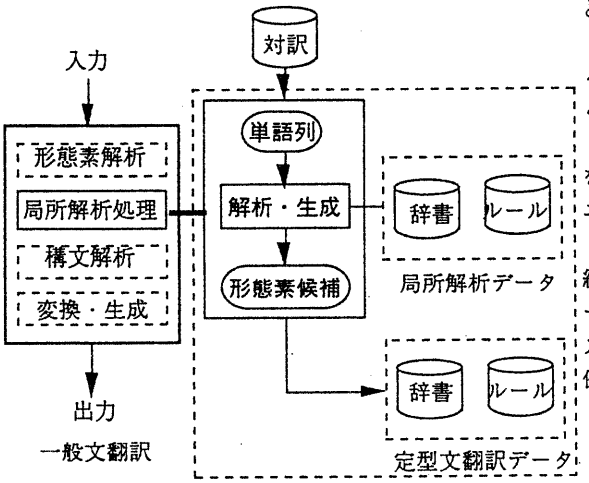


図8　定型文翻訳データ自動作成システム

3．4　自動作成上の問題

今までは，数量表現は局所解析処理で得られた日本語訳と，人手で与えた対訳中の日本語訳が一致するものとしてきた．しかし，例えば，例文1の数量表現 "17.76 dollars per kilo" は

1）「1キロにつき17.76 ドル」

2）「17.76 ドル／キロ」

と翻訳することも可能であり，この場合，自動作成では数量表現が変数とはならない．

この問題を避けるために，次のように局所解析ルールを変更する．

英語解析ルール	日本語訳
S --> UNTEXP	「#1#」
--> UNTEXP1	「#1#」
--> UNTEXP2	「#1#」
--> UNTEXP3	「#1#」
UNTEXP --> NUMEXP UNIT	「#1##2#」
UNTEXP1 --> UNTEXP PER UNIT	「1#3##1#」
UNTEXP2 --> UNTEXP PER UNIT	「1#3#につき#1#」
UNTEXP3 --> UNTEXP PER UNIT	「#1#／#3#」

図9　修正された局所解析データ

図9では，同じ英語表現に対して，異なる日本語訳がある場合には，それぞれの日本語訳の数だけ英語解析ルールを分けた．そのため，ルール数が増えるという欠点はあるが，1），2）の場合とも自動作成できるようになる．

このとき，1），2）の場合に自動作成で得られる英語解析ルールは，それぞれ

1) S --> PAT1 CMA PAT2 UNTEXP2 CMA UPDW UNTEXP

2) S --> PAT1 CMA PAT2 UNTEXP3 CMA UPDW UNTEXP

となる．

4．外電経済ニュース翻訳システム

4．1　システム概要

定型文を多く含む外電経済ニュースに翻訳対象を絞り，定型文翻訳方法を組み込んだ外電経済ニュース翻訳システムについて述べる．

ＡＰ電は一日３５０ほどのニュースが入電し，経済ニュースはそのうち５０ほどである．各ニュースには必ずタイトルがついており，そのニュースの特徴を表している．図10に外電のニュースの例を示す．

```
W1001u FBX  TXB803  24-04 00039                    ヘッダー
87u 01 233u psm ldm

^^Yen-Dollar Opening                               タイトル

  TOKYO (AP) - The U.S. dollar opened at 137.50 yen on
the Tokyo foreign exchange marke .....
                                                   本文
  END

AP-TK-24-04-91 0004GMT<                            トレイラー
```

図10　ＡＰ電ニュース

"^^Yen-Dollar Opening"がタイトルであり，"The U.S. dollar opened..."からニュース本文が始まる．実際にはこのようなニュースが連続して入ってくる．また，本文が表の形式をした表ニュースも多い．

ここでタイトルに注目すると，表ニュースでは

例えば

「ゴルフの結果」,「株の取引価格」

等，経済ニュースでは

「金価格」，「日本市場」

等，毎日固定したものが使われる．したがって，タイトルによって自動的に表ニュースや経済ニュースを選別することができる．

図11に外電経済ニュース翻訳システムの処理の流れを示す．

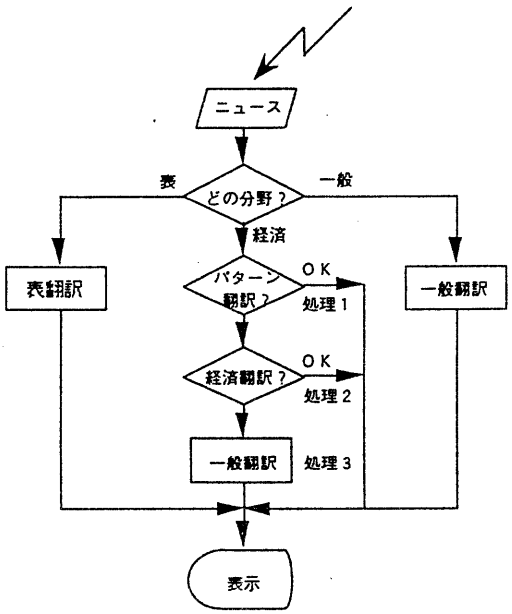


図11　外電経済ニュース翻訳システム

本システムでは始めにタイトルによって，表ニュース，経済ニュース，一般ニュースの３種類にニュースを分類する．表ニュースは表翻訳ルーチンに進み，英単語を日本語に一対一に単純に置き換える処理のみが行なわれる．（今回は，表ニュースの中でも経済に関係するものしか扱わなかった．）一般ニュースは従来からの機械翻訳システムで処理される．

経済ニュースに関しては以下で説明する３つの処理を行ない，いずれかの処理で翻訳された場合には，結果を出力し以降の処理を中止する．

処理１は，定型文を翻訳する．すなわち，主に数量表現や時間表現のみが変化するような文のみを翻訳の対象としており，形容詞や副詞などが１つでも付加されたような文は翻訳できない．そこで処理２では若干の柔軟性を持たせるために，定型文に完全に一致しない場合でも翻訳できるように文法や辞書を作成した．例えば，図12に示す例のように，"Tuesday morning in Tokyo","Wednesday morning in Tokyo","in Tokyo Thursday"等，大まかな構文が非常に似た文ではあるが，使われている単語や順番が若干異なる部分を持つ文を対象にしている．

The U.S. dollar opened *slightly higher* against the Japanese yen Tuesday morning in Tokyo, **while** share prices *inched up*.

The U.S. dollar opened *lower* against the Japanese yen Wednesday morning in Tokyo, **while** share prices *slipped*.

The U.S. dollar opened *higher* against the Japanese yen in Tokyo Thursday, as share prices *rose in early trading*.

図12　処理２で翻訳される文の例

これらいずれの処理でも翻訳されなかった文は，処理３で一般の文を処理する過程と全く同じ翻訳が行なわれる．

４．２　各翻訳処理

各翻訳処理の諸元を表１に示す．以下，各処理について具体的に説明する．

表１　各翻訳処理の諸元

	翻訳方式	文法	辞書	処理速度
処理1	解析：パターンマッチング 生成：代入	対訳パターン 約400 局所解析用 約60	1,400 (局所解析用) 1,900 (対訳データ)	5秒／文
処理2	トランスファー方式	約500	基本語 57,000 経済語 31,000	10秒／文
処理3	トランスファー方式	約2,500	基本語 57,000 専門語 111,000	20秒／文

４．２．１　処理１の諸元

２．で得られた定型文中から数字，曜日等の同一化を行なった後，経済ニュースに関する異なり文上位４００文を人手で抽出した．それらに対して日本語訳を人手で与え，３．で述べた定型文自動登録方法により，定型文翻訳データを自動的に作成した．ほとんどのサンプル定型文が数量表現を含んでいたが，

Gold prices were mixed

「金価格は小動きだった。」

のように変数を全く含まない文もあった．この場

合には，英語解析ルールは，

S --> PAT225

PAT225 --> "Gold prices were mixed"

となる．

4．2．2　処理2の諸元[4]

ここで使用する文法ルールと辞書は，'91年4月23日～25日のＡＰ電３日分（約４５０文が含まれている）を用いて作成した．経済ニュース用に文法ルールと辞書をチューニングするために使った，この３日分のデータを，以下では学習データと呼ぶ．解析できる文の種類を少なくすることによりルール数を少なくすることができ，またルール上で前置詞句の係り先もかなり特定できたので，構文的曖昧性を抑えることができた．しかし，システムの制約上，このすべての文を翻訳することができるように文法ルールや辞書を作成できたわけではない．

4．2．3　処理3の諸元

処理３は従来の翻訳システムである．辞書は，処理２で使用した経済専門辞書の名詞のみを含む大規模のものを用い，文法ルールも一般のものである．一般文も処理３と同様に処理される．

5．評価

学習データと非学習データ（'91年３月７日のＡＰ電，約１６０文）に対する処理過程別処理率と翻訳率を表２に示す．ここで評価にはMuプロジェクトの方法を用い，評価値上位２つ，すなわち，

１．入力文の文意は出力文に忠実に再現．

２．文意は再現されているが，細部に問題．

を正解とした．

表２　処理過程別処理率と翻訳率（単位%）

		処理１	処理２	処理３	TOTAL
学習	処理率	29.1	61.3	9.6	100
	翻訳率	100	70.1	10.2	73.0
未学習	処理率	31.2	57.5	11.3	100
	翻訳率	100	58.7	22.2	66.3

学習データでは約３０%の文が処理１で翻訳され，約６０%の文が処理２で翻訳されている．翻訳率は処理１では当然のことながら１００%，処理２では約７０%である．全体の翻訳率は約７３%と高い．

学習データでも非学習データでも各処理における処理率はほとんど同じである．非学習データでは学習データに比べると処理２の翻訳率が１０%低くなっているのに伴い，全体の翻訳率も約１０%下がっているものの，従来の翻訳率に比べ格段に向上している．処理３に残った文は非常に長いものが多かった．

6．おわりに

定型文の抽出とその機械翻訳について述べた．また，定型文翻訳を組み込んだ，外電経済ニュース機械翻訳システムを作成し，その翻訳率の評価を行なった．経済ニュースは特有な文型を持っている場合が多く，その特徴を使うことによって従来より，翻訳率が格段に上がった．定型文翻訳処理では，パターンが登録されているかどうかは早い段階で決まるので，ここでの処理時間は非常に少ない．したがって，全体としてもそれほど翻訳時間はかからない．

今後は，定型文抽出に関してはパラメータをいろいろ変え，抽出される文の比較検討を行ないたい．また，定型文翻訳をスポーツニュース，一般のニュースに拡大できないかどうかを検討し，外電ニュース全体の翻訳率向上をめざす．

［参考文献］

[1]浦谷，加藤，相沢「ＡＰ電経済ニュースからの定型パターンの抽出」情報処理学会第42回全国大会(1991)

[2]加藤，浦谷，相沢，中瀬「英日機械翻訳における固有名詞処理」情報処理学会第40回全国大会(1990)

[3]野村「自然言語処理の基礎技術」電子情報通信学会(1988)

[4]相沢，鎌田「ＡＰ電経済ニュースの英語解析用文法」情報処理学会第45回全国大会(1992)

［付録］定型文の例（含有率；定型文）

1.000;He did not elaborate.

1.000;No injuries were reported.

1.000;The U.S. dollar opened at 159.97 yen on the Tokyo foreign exchange market Monday, up from last Friday's close of 157.65 yen.

0.970;The average price for strict low middling 1 1-16 inch spot cotton declined 99 points to 78.64 cents a pound Wednesday for the seven markets, according to the New York Cotton Exchange.

0.852;Philippine peso banknotes Friday at 20.50-21.00 pesos (dealer buying-dealer selling) per U.S. dollar at the close, unchanged from a day earlier.