# 語の接続関係を利用した機械翻訳システム†

鈴木康広†† 栃内香次††

現在一般に用いられている機械翻訳の手法は，原言語の構文解析，意味解析を行った後，その情報に基づいて目的言語に変換する方式である．我々はこれとは異なる手法として語の接続関係を機械翻訳に応用する研究を行っている．ある限られた分野の文献を対象とする場合，文章を構成するそれぞれの単語の間には特定の接続関係がある．例えば，「処理，に関する，言語，研究，自然」という語群がある場合，これらの語を並べて生成することができる文章のうち，最も自然な文章は「自然言語処理に関する研究」というようにほぼ1文に限定することができる．すなわち，ある単語に接続可能な単語は少数に限定されていると考えられる．したがって，ある分野の学術文献などから文章中で接続している2つの語を接続情報として大量に抽出してあらかじめ辞書に登録しておき，英文を単語単位に翻訳して得られた訳語群から接続情報を用いて一意に翻訳文を作成することができる．実際に，実験システムを作成し情報処理関係の論文などの表題116例について翻訳実験を行った結果，英日翻訳で75%の正翻訳率が得られ，日英翻訳で66%の正翻訳率が得られた．さらに，情報処理関係の論文などの一般文についても翻訳実験を行った結果，このアルゴリズムが一般文の翻訳についても適用可能であることがわかった．

## 1. はじめに

現在一般的な機械翻訳の手法は，原言語の構文解析を行った後，その情報に基づいて目的言語に変換する方式である[1]．本論文では，これとは別の手法として，語の接続関係を用いた機械翻訳手法を提案する[2]．この手法は，辞書中に品詞，活用などの文法的な情報を登録する必要がなく，単語情報の辞書への登録が容易であるという特長を持っている．

ある限られた分野の文献を対象とする場合，文章を構成するそれぞれの単語の間には特定の接続関係がある[3]．例えば，「処理，に関する，言語，研究，自然」という単語群が存在する場合，これらの語を並べて作ることができる文章のうち，最も自然な文章は「自然言語処理に関する研究」というようにほぼ1文に限定される．すなわち，ある単語に接続可能な単語は少数に限定されている．そして，対象文章の分野を限定することによりこの性質はさらに著しくなると予想される．したがって，ある特定分野の多数の学術文献から，文章中で接続しているそれぞれの単語の組をあらかじめ抽出して辞書に登録しておくことによって，文章を構成する単語群が与えられた場合，上記の情報をもとに文章を生成することができる．本論文では，このような語の接続関係を利用した機械翻訳手法の可能性について述べるものである．

† Machine Translation System Using Conjunctive Relations of Words by YASUHIRO SUZUKI and KOJI TOCHINAI (Department of Electronic Engineering, Faculty of Engineering, Hokkaido University).

†† 北海道大学工学部電子工学科

## 2. 翻訳アルゴリズム

前述のように，文章中のある単語について，その単語に接続可能な単語は特定の単語に限ることができる．このことを利用して，以下に述べるような機械翻訳のアルゴリズムが考えられる．

前述の，「処理，に関する，言語，研究，自然」という語群を考えてみる．この語群から生成することのできる文章は，「自然言語処理に関する研究」という一文に定まる．これは，それぞれの単語が「T—自然，自然—言語，言語—処理，処理—に関する，に関する—研究，研究—E（Tは文頭，Eは文末を意味する）」という接続関係を持ち，文はこのような接続関係の連鎖で表されることを示している．この関係を接続情報と呼ぶことにする．そこで，次の例に示すように英文を構成する各単語を単語単位に翻訳して得られた語群から，接続情報を用いて翻訳文を作成することができる．

例)

Study on Natural Language Processing
↓ ↓ ↓ ↓ ↓ (英文)
研究 に関する 自然 言語 処理
(単語単位の翻訳)

T—自然，自然—言語，言語—処理，処理—に関する，に関する—研究，研究—E
(接続情報)

自然 言語 処理 に関する 研究 (訳文)

ここで，接続情報はあらかじめ同一分野の大量の文献（文数にして1500～2000文程度）から抽出し，接続情報辞書に蓄えておくものとする．また，このアルゴリ

ズムは日英翻訳についても適用可能である．すなわち，英文についても同様に接続情報，「T—Study, Study—on, on—Natural, Natural—Language, Language—Processing, Processing—E」をあらかじめ辞書に登録しておくことによって以下の例に示すように翻訳が可能である．

例)

自然 言語 処理 に関する 研究
↓ ↓ ↓ ↓ ↓

(日本語)

Natural Language Processing on Study

(単語単位の翻訳)

T—Study, Study—on, on—Natural
Natural—Language, Language—Processing
Processing—E

(接続情報)

Study on Natural Language Processing

(訳文)

すなわち，本手法は双方向に適用可能である．また，多義語における訳語の選択も接続情報を用いることによって行うことができる．

## 3. 実験システム

上述のアルゴリズムにもとづく実験システムを作成した．以下，このシステムの概要と翻訳手順について述べる．実験システムは PL/I で書かれており，北海道大学大型計算機センタの HITAC M-680 H 上に作成されている．翻訳の手順は基本的には英日と日英翻訳で共通である．なお，本システムは小規模な実験システムであり，翻訳対象文章は主として情報処理関係の論文，研究会報告等の表題に限定している．実験システムの処理の流れを図1に示す．ここで，単語辞書および接続情報辞書にはあらかじめ情報処理学会論文誌および全国大会予稿の表題 1500 例から抽出した情報を登録してある．登録された単語数は 1463 語，日本語接続情報数は 5317 組，英語接続情報数は 5747 組である．以下，図1に従って翻訳処理の概略を述べる．

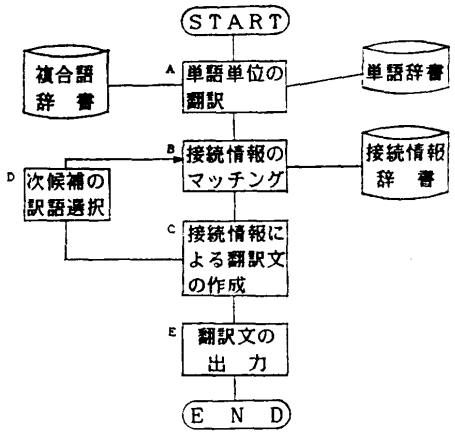



図 1 翻訳処理の流れ
Fig. 1 Flow of transaction.

### 3.1 英日翻訳処理

A．単語単位の翻訳

単語辞書を用いて翻訳対象文章の単語単位の翻訳を行う．単語辞書には，英単語とそれに対する日本語の訳語が3種類まで記録されている．単語辞書の構造を図2に示す．この辞書から最も頻度の大きい訳語を選択して，単語単位の翻訳を行う．以下に "A Study on Natural Language Processing" の単語単位の翻訳例を示す．

| GN1 | GN2 | WORD | $M_1$ | $H_1$ | $M_2$ | $H_2$ | $M_3$ | $H_3$ | CP | FR |
|---|---|---|---|---|---|---|---|---|---|---|
| 134 | 345 | analysis | ｶｲｾｷ | 73 | ﾌﾞﾝｾｷ | 2 | | 02 | | 75 |
| 165 | | automatic | ｼﾞﾄﾞｳ | 41 | | 0 | | 01 | | 41 |
| 135 | 431 | computer | ｺﾝﾋﾟｭｰﾀ | 42 | ｹｲｲﾝｷ | 38 | | 02 | | 80 |

GN1 ：語番号
GN2 ：この語の類似語番号（単数，複数形）
WORD：英単語
$M_i$ ：訳語
$H_i$ ：訳語の頻度
CP ：登録されている訳語の数
FR ：英単語の出現頻度

図 2 単語辞書の構造
Fig. 2 Structure of word's dictionary.

表 1 複合語辞書の内容
Table 1 Content of compound word's dictionary.

| 英語複合語 | 日本語複合語 |
|---|---|
| Production Rule | プロダクション ルール |
| Expert System | エキスパート システム |
| Machine Translation | 機械 翻訳 |
| Personal Computer | パーソナル コンピュータ |
| Kana-kanji Translation | かな漢字 変換 |
| Natural Language | 自然 言語 |
| Production System | プロダクション システム |
| Image Processing | 画像 処理 |
| Computer Graphic | コンピュータ グラフィック |
| User Interface | ユーザ インタフェース |
| Vector Processor | ベクトル プロセッサ |
| Word Processor | ワード プロセッサ |
| Parallel Computer | 並列 計算機 |

A Study on Natural Language Processing
↓ ↓ ↓ ↓ ↓
研究 に関する 自然 言語 処理

単語単位の翻訳を行う際，ある単語が複数の訳語を持ち，かつほぼ等頻度で出現している場合は頻度情報による訳語の選択は困難である．このような場合，当該の語がその前後の語と複合語になっていれば，それを利用して訳語を選択することができる．すなわち，単語辞書とは別に複合語辞書を設け，当該の語が複合語辞書に登録されている場合は無条件で複合語辞書中の訳語を選択する．なお，単語単位の翻訳の際，複合語の処理を優先させる．辞書に登録されている複合語の一部を表1に示す．また，'call-on', 'consist-of' などは，一つの単語として「要求する」,「からなる」というように単語辞書に登録しておく．

B．接続情報のマッチング

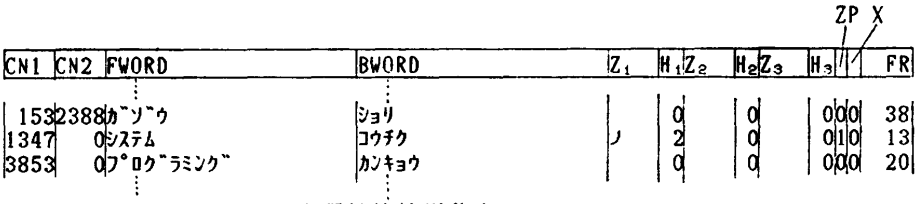
単語単位の翻訳によって得られたそれぞれの訳語について，日本語接続情報辞書を用いて接続可能な訳語の組を捜す．以下に示す例は "Study on Natural Language Processing" を単語単位に翻訳した結果から得られる接続情報である．

例) T—自然，T—言語，T—処理，研究—E
に関する—研究，自然—言語，言語—処理
言語—E，処理—言語，処理—E
処理—に関する

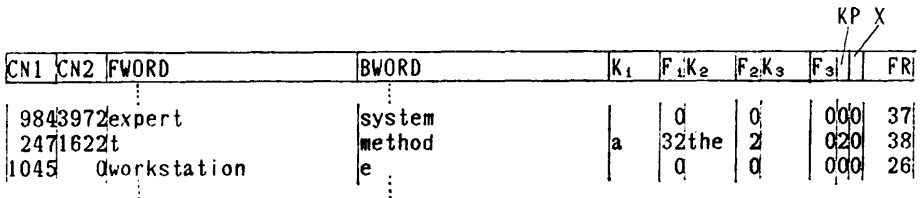
日本語接続情報辞書の構造を図3-a に示す．図に示すようにこの辞書には訳語の組，すなわち接続情報とその出現頻度および訳語間に入る可能性のある助詞が3種類までその出現頻度とともに記録されている．

C．接続情報による翻訳文の作成

ステップBで求めた訳語の組を用いて文章の組み立




a 日本語接続情報辞書

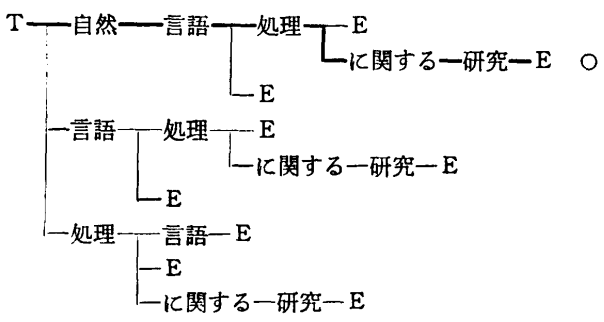



b 英語接続情報辞書

CN1 ：接続情報番号　CN2 ：類似接続情報番号
FWORD：前の単語　BWORD：後の単語
$Z_i$ ：単語間の助詞　$H_i$ ：助詞の出現頻度
$K_i$ ：単語間の冠詞　$F_i$ ：冠詞の出現頻度
ZP ：登録されている助詞の数　KP ：登録されている冠詞の数
X ：未使用　FR ：接続情報の出現頻度
*) FWORD, BWORD 中の t, e は文頭，文末を表す

図 3 接続情報辞書の構造
Fig. 3 Structure of conjunctive information's dictionary.

てを行い，すべての訳語が含まれ，かつ文頭から文末まで接続している訳語列を訳文とする．以下に，"Study on Natural Language Processing" に対する接続可能な訳語列を示す．




○

ここで，○印の付いている文章が訳文となる．

翻訳文の生成過程で辞書中に接続情報が存在しないために訳文が生成されない場合がある．このような場合は，以下のルールを用いて訳文を生成する．

〈ルール〉　1か所または2か所で接続情報が辞書中に存在しない場合は，その部分を強制的に接続し訳文とする．

例1）「接続—関係」が辞書にない．

T—語—の—接続
⇔関係—を用いた—機械—翻訳—E

例2）「接続—関係」，「を用いた—機械」が辞書にない．

T—語—の—接続⇔関係—を用いた
⇔機械—翻訳—E

上の例で，例1は一つの接続情報「接続—関係」が接続情報辞書中にない場合で，例2は二つの接続情報「接続—関係」，「を用いた—機械」が接続情報辞書にない場合であり，'⇔' は強制的に接続した部分である．なお，上述のルールは日英翻訳時にも同様に適用される．

D．次候補の訳語選択

3か所以上で接続情報が辞書中に存在しない場合はステップCの処理で訳文が生成されない．このとき，以下の手順で訳語の次候補を選択する．

① 今，単語単位の翻訳の結果 $A_1, B_1, C_1, D_1, E_1, F_1$ という訳語が得られたとする．このうち次候補を持つものを $A_1, E_1$ とし，これらの次候補を $A_2, E_2$ とすると，以下のように次候補選択が行われる．

イ）次候補を持つ訳語 $A_1, E_1$ についてこれらを含む接続情報をステップBで得られた接続情報から捜す．

ロ）$A_1, E_1$ を含む接続情報の出現頻度の総和

$S_A = \{\sum(A_1-X)+\sum(X-A_1)\}$

$S_E = \{\sum(E_1-X)+\sum)(X-E_1)\}$

を求める．ここで，$(A_1-X)$, $(X-A_1)$ は訳語 $A_1$ を含む任意の接続情報の出現頻度であり，$\sum(A_1-X)$, $\sum(X-A_1)$ は接続情報の出現頻度の総和である．

ハ）$S_A, S_E$ のうち最小のものについて次候補を選択する．

$S_A < S_E$→次候補 $A_2$ を選択

$S_A > S_E$→次候補 $E_2$ を選択

ただし，複数の訳語を持つ語が一語しか存在しない場合は，その訳語の次候補を選択する．

② ①で取り出した次候補を用いて再び接続情報のマッチングを行う．

③ 訳文が生成されない場合は，候補がなくなるまで①，②を繰り返す．

④ 最終的に訳文が生成されなかった場合は，解析不能のメッセージを表示する．

E．翻訳文の出力

生成された文章が一文の場合はその文章を翻訳文として出力する．生成された文章が複数の場合は以下の手順で翻訳文を選択する．

① それぞれの接続情報には，頻度情報が付加されているので文頭から文末までの各接続情報の頻度の総和を求める．

② ①で求めた接続情報の頻度の総和が最大の文章を翻訳文として選択する．

②では，接続情報の頻度の総和が最大である翻訳結果は，過去にその単語の並びで出現した回数が一番多いことになり，正しく翻訳される可能性も高いことからこのような処理を行っている．

### 3.2 日英翻訳処理

A．単語単位の翻訳

日英翻訳の場合は逆に単語辞書の訳語の方をキーとして最も頻度の高い英単語を選択し，単語単位の翻訳を行う．以下に「自然言語処理に関する研究」の単語単位の翻訳例を示す．

自然　言語　処理　に関する　研究
↓　↓　↓　↓　↓
Natural Language Processing　on　Study

単語単位の翻訳を行う際，英日翻訳と同様に複合語辞書中の訳語を優先させる．

B．接続情報のマッチング

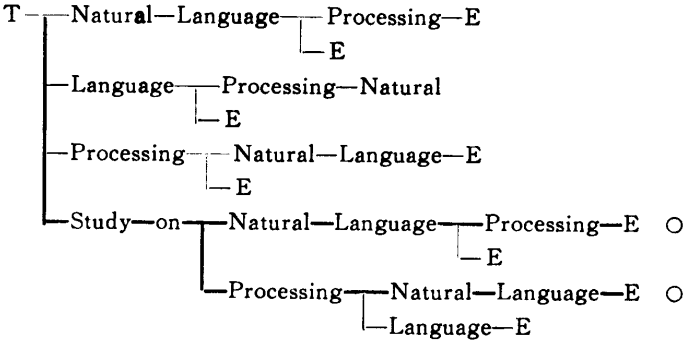
英語接続情報辞書を用いて接続可能な訳語の組を捜す．以下は上記の結果から得られる接続情報である．

T—Natural, T—Language, T—Processing
T—Study, Natural—Language, Study—E
Language—Processing, Language—E
Processing—Natural, Processing—Language
Processing—E, On—Natural, On—Processing

英語接続情報辞書の構造を図3-b に示す．図に示すようにこの辞書には接続情報とその出現頻度および後の訳語に付随する冠詞が3種類 (a, the, なし)，その出現頻度とともに記録されている．なお，英語接続情報辞書は日本語接続情報辞書とともにあらかじめ人手によって作成しておく．

C．接続情報による翻訳文の作成

ステップBで得られた接続情報を用いて訳文を生成する．以下に「自然言語処理に関する研究」の訳文生成過程を示す．




この例では○印を付した2文が訳文として生成される．

D．次候補の訳語選択

英日翻訳の場合と同様な方法で次候補の訳語を選択する．

E．翻訳文の出力

ステップCで生成された訳文を以下に示す．

① A Study on Natural Language Processing (124)
② A Study on Processing Natural Language (116)

上の例で（ ）内は接続情報の頻度の総和である．上の例では①の訳文が最終的な翻訳結果として選択される．

### 3.3 日英翻訳における冠詞の処理

日英翻訳時には冠詞の処理の問題が生じる．本システムでは冠詞の処理を統計的に行っている．以下に，冠詞の処理手順を示す．

① 単語単位の翻訳結果から英語接続情報辞書によって得られた訳語の組に対して，訳語の組の出現頻度，および冠詞とその出現頻度を辞書から取り出す．

② ①で取り出した訳語の組の出現頻度を $M$，冠詞の出現頻度（複数の冠詞が付属する場合はそれぞれの冠詞の出現頻度の総和）を $N$ とし

$(N/M)\times 100$ [%]

の値が 50% を越える場合，最も出現頻度の大きい冠詞を付ける．

③ 50% 以下の場合は冠詞を付けない．

上述の 50% という値は以下に示すように実験的に定めたものである．すなわち，接続情報の出現頻度に対する冠詞の出現頻度の割合をパラメータとして冠詞の当てはめ実験を行ったところ**表2**に示すような結果が得られた．表2は冠詞がつく可能性のある267か所について冠詞の当てはめを行った結果であり，PARA は冠詞を付けるか付けないかの判断の基準となる前述のパラメータ $N/M$ である．正解率は267か所中原文と一致した割合を示す．a，b，c は誤りの種類で，それぞれ原文に冠詞が付いているのに付けなかった誤り（a），原文に冠詞が付いていないのに付けた誤り（b），誤った冠詞を付けた誤り（c）である．表2をみると PARA＝50% のときが最良の値となっている．

表 2 冠詞の当てはめ実験の結果
Table 2 Results of the experiment. (%)

| PARA | 正解率 | a | b | c |
|---|---|---|---|---|
| 10 | 57.3 | 8.8 | 86.9 | 5.2 |
| 30 | 67.0 | 44.3 | 53.4 | 2.3 |
| 50 | 72.7 | 94.7 | 4.0 | 1.3 |
| 70 | 70.8 | 100 | 0 | 0 |
| 90 | 64.8 | 100 | 0 | 0 |

### 3.4 英日翻訳における助詞の処理

英日翻訳時には，「は」，「が」などの助詞の処理の問題が生じる．これについて冠詞の場合と同様の処理を行っている．ただし，接続情報の延べ出現数に対して助詞の付随する割合は 0.9% で極めて小さく，PARA の値を実験的に決定することはできなかったので，冠詞の場合を参考にして 50% としている．

## 4. 翻訳実験と実験結果

### 4.1 論文表題の翻訳実験

実験システムの性能を評価するために，情報処理関

係の文献の表題116例について英日，日英の翻訳実験を行った．翻訳実験の手順を以下に示す．

1. 論文表題 1500 例から情報を抽出し，単語辞書，日本語接続情報辞書，英語接続情報辞書をあらかじめ作成する．
2. 辞書作成に用いた論文表題とは別の 116 例の表題について英日，日英翻訳を試みる．
3. 2. で新出語，新出接続情報が出現し翻訳不能となった場合はこれらを辞書に登録する．

なお，実験に用いた表題には著者が書いた日本語表題と英語表題があり，翻訳結果は原則として原文と一致した訳文が得られた場合を正しく翻訳されたと見なした．ただし，以下に示すような冠詞の違いおよび単数，複数の違いや表記上の揺れがあった場合でも正しく翻訳されたと見なしている．

例) A Method～＝Method～

～Expert Systems～＝～Expert System～

～を用いた～＝～を利用した～

〈英日翻訳実験結果〉

116例中100例（86%）が翻訳可能であった．翻訳不可能であった 16 例のうち接続情報が辞書中に存在せず，次候補選択を行ってもなお解析不能であった例が 12 例，単語が辞書中に存在しなかった例が4例であった．翻訳可能であった100例のうち正しく翻訳された例が 87 例（87%）であり，最終的な正翻訳率は 75% であった．また，100 例中出力結果が1文であった例が 58 例（58%）で，その正翻訳率は 95% であり，出力結果が複数であった例が 42 例（42%）で，その正翻訳率は 76% であった．なお，複数例が出力された場合の平均文数は6.3文であった．

〈日英翻訳実験結果〉

116 例中 98 例（85%）が翻訳可能であった．翻訳不可能であった 18 例のうち接続情報が辞書中に存在せず，次候補選択を行ってもなお解析不能であった例が 14 例，単語が辞書中に存在しなかった例が4例であった．翻訳可能であった 98 例のうち正しく翻訳された例が 76 例（78%）であり，最終的な正翻訳率は 66% であった．また，98 例中出力結果が1文であった例が 48 例（49%）で，その正翻訳率は 90% であり，出力結果が複数であった例が 50 例（51%）で，その正翻訳率は 66% であった．なお，複数例が出力された場合の平均文数は7.6文であった．

なお，翻訳可能な場合とは文頭から文末まで接続情報によって単語が接続し，システムが1文以上の翻訳結果を出力した場合を言う．

表 3 複数文出力における正解含有率
Table 3 The content rate of exactly translated sentences.

| | 英日翻訳 | 日英翻訳 |
|---|---|---|
| | 正解含有率 (%) | 正解含有率 (%) |
| ① | 76.2 | 66.0 |
| ② | 83.3 | 70.0 |
| ③ | 83.3 | 88.0 |

① 評価順位1位の文が正解である割合
② 評価順位2位以上の文の中に正解が含まれる割合
③ 出力結果の中に正解が含まれている割合

翻訳結果をみると，最終的な正翻訳率は日英翻訳の方が約 10% 低い値となっている．この原因としては，辞書構造の関係で英語と日本語単語の多義性に違いが生じたためと考えられる．英単語の場合は，辞書の構造から最大3種類の訳語までしか持つことができないのに対して，日本語単語の場合は，同じ辞書を用いているので制限がない．このことによって日英翻訳の場合，その接続情報にばらつきが生じ，正しい翻訳ができず正翻訳率が低くなっていると考えられる．また，もう一つの原因として日本語と英語の構文上の違いが考えられる．一般に英語の方が語順などが重要で接続関係が厳しく制限される．このことが日英翻訳に影響を及ぼしていると考えられる．今回の実験では単語辞書を英日，日英翻訳で共用しているが，日英翻訳のための単語辞書を別に作成して訳語の個数に制限を設ければ正翻訳率を改善することができると考えられる．

また，一般的に複数の訳文が出力されたときの正翻訳率が低くなっている．これは，複数の訳文が作成された場合，接続情報の頻度の総和が最大のものを適訳として選択しているからである．**表3**は，複数文が出力されたときの正解含有率を示す表であり，それぞれの訳文の接続情報の頻度の総和を求め，頻度の総和の大きい順に順位を付け，1位のものが正解である割合，2位以上のものが正解である割合，出力結果の中に正解が含まれている割合をそれぞれ示したものである．表3を見ると評価順位が2位以上の文に正解が含まれている割合がかなり高いので，評価順位1位の文だけを適訳として選択するのではなく，評価順位2位の文を次候補として出力し，人間の選択に任す方が良いと考えられる．

また，複数の訳文が出力された場合で，接続情報の頻度の総和が最大であるのに不正解であった例を調査

表 4 一般文の翻訳実験の結果
Table 4 Results of the experiment.

| 入力文数 | 50 | 100 | 150 | 200 | 250 | 300 | 350 |
|---|---|---|---|---|---|---|---|
| 正翻訳率 (%) | 17 | 20 | 34 | 30 | 32 | 36 | 38 |

すると，以下に示すように意味的にはほぼ等しい例があることがわかった．このような文が適訳として出力された場合を正解とすると英日および日英の正翻訳率は 3〜4% 高くなる．

例) A Support Tool for Production System Development

プロダクションシステム開発のための支援ツール (正解)

プロダクションシステム開発支援のためのツール

### 4.2 一般文の翻訳実験

はじめに述べたように，本システムは論文表題を対象としている．しかし，原理的には一般の文に対しても同一のアルゴリズムが適用可能であると考えられる．そこで次に，一般文を対象として翻訳実験を行った．実験は表題の翻訳実験に使用した辞書を初期辞書として用い，辞書は一文ごとに新出語登録を行って更新した．なお，対象文章は情報処理関係の英文論文誌 44 編 (画像処理，VLSI 設計，並列処理，信号処理等) のアブストラクトおよび序論に出現する文章で単文のみを対象とした．翻訳実験の結果をまとめて**表 4** および**図 4** に示す．表 4 は，一般文の翻訳実験結果であり，一文ごとに辞書を更新し，50 文ごとに正翻訳率とその推移を求めたものである．図 4 を見ると一般文 300 文から情報を抽出した時点で正翻訳率が 35% 程度となっている．この値は，まだ辞書の内容が充実していないので低い値となっているが 1000 文程度から単語および接続情報を抽出することにより表題の場合と同程度の約 70% の正翻訳率を得ることができると考えられる．なお，300 文について翻訳を行った結果，翻訳結果が得られた文は 89 文であり，この 89 文のうち 83 文が正解であった．翻訳不能であった文について調べると，ほとんどの文が辞書中に単語あるいは接続情報が存在せず翻訳不能となっている．また，解析可能であった 89 文のうち 7 文について複数の訳文が出力され，これらから適訳として選ばれた 7 文のうち 6 文が誤りであった．複数の訳文が出力された例をいくつか示す．

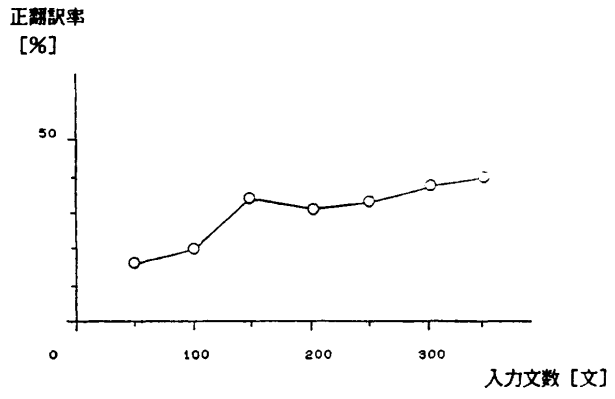



図 4 一般文における正翻訳率の推移
Fig. 4 Change in the rate of exactly translated sentences.

※ This paper discusses a VLSI algorithm for pattern matching.

① 本論文は VLSI のためのパターンマッチングアルゴリズムについて論じる. (163)

② 本論文はパターンマッチングのための VLSI アルゴリズムについて論じる. (正解) (149)

※ For this problem we present an efficient algorithm.

① 我々はこの問題のために能率的アルゴリズムを示す. (61)

② この問題のために我々は能率的アルゴリズムを示す. (正解) (61)

上の例を見ると，文章全体の意味は①と②で等しい．原文と一致しているのはいずれも②の文である．一般文の場合も論文表題と同じく評価順位 1 位の文のみを適訳として選択するのではなく評価順位 2 位の文も次候補として選択する方が良いと考えられる．

今後の課題としては，

1. 一般文の接続情報の増大による翻訳可能率の改善，
2. 複数の翻訳結果からの適訳の選択アルゴリズムの検討.

の 2 点が挙げられる．1. については，表題の場合は 1500 例から情報を抽出した時点で辞書の内容が充実し，70% 程度の正翻訳率が得られている．また，図 4 を見ると正翻訳率は最初の部分を除き，入力文数の増加に従ってほぼ直線的に増加する傾向を示している．したがって，これらのことから一般文の場合も 1500 文程度の入力により辞書の内容が充実し，表題文の場合と同程度の正翻訳率が得られると予想される．しかし，一般文は表題文に比べ文の構造がより多様になるので，これを確証するためには一般文について 1000〜

1500 文程度まで入力文数を増して翻訳実験を行い，正翻訳率の推移を求める実験を行う必要があると考えられる．2. については，辞書中の接続情報が増加すると複数の翻訳結果が作成される場合が多くなるので，複数の翻訳結果から適訳を選択するアルゴリズムの検討が今後の課題となる．

最後に一般文の翻訳例をいくつか示す．

〈英日翻訳〉

・An action routine description is a set of fragments of programs associated-with production rules.

→動作ルーチン記述はプロダクションルールに付随するプログラムの断片の集合である．

・This paper describes the design of a software development system.

→本論文はソフトウェア開発システムの設計について述べる．

〈日英翻訳〉

・本論文では，組み合わせ問題向きマルチコンピュータシステムが提案されている．

→In-this-paper, a combinatorial problem oriented multicomputer system is-proposed.

・プログラムのそれらの断片は上昇型構文解析系によって与えられる順序に従って起動される．

→Those fragments of programs are-activated according-to the ordering given by a bottom-up syntax analyzer.

## 5. おわりに

本論文では，語の接続関係を利用した機械翻訳手法について述べてきた．本手法は，文法的な情報を用いず接続情報のマッチングで訳文を生成するので以下のような利点がある．

1. 辞書への単語情報の登録が容易である．
2. システムでの実現が容易である．

実際に，実験システムを作成し情報処理関係の文献の題目を対象とした翻訳実験を行った結果，英日翻訳で 75%，日英翻訳で 66% の正翻訳率が得られた．また，このアルゴリズムのままで一般文についてもある程度の翻訳が可能であることがわかった．

今後の課題としては，複数の訳文が出力された場合の適訳の選択や一般文の翻訳のための接続情報辞書の充実および複文の処理などを検討するとともに従来の手法との比較を行っていく予定である．また，システムとして実現する場合，辞書の更新をいつどのように行うかという問題も今後の課題である．

**謝辞** 本研究を行うにあたり，終始適切な御示唆をいただいた本学部電子工学科電子機器工学講座各位に感謝します．

## 参考文献

1) 情報処理「特集：機械翻訳」，Vol. 26, No. 10 (1985).
2) 鈴木ほか：語の接続関係を用いた機械翻訳，第35回情報処理学会全国大会論文集，3S-3 (1987).
3) 鈴木：日本語情報処理における語の接続関係とその応用に関する研究，北海道大学工学部修士論文 (1985).

(昭和62年9月 9 日受付)
(昭和63年1月19日採録)

**鈴木 康広**（正会員）

昭和35年生．昭和57年北海道工業大学電気工学科卒業．昭和60年北海道大学大学院工学研究科修士課程情報工学専攻修了．現在同大学大学院博士後期課程在学中．語の接続関係を利用した日本語情報処理の研究に従事．電子情報通信学会，IEEE 各会員．

**栃内 香次**（正会員）

昭和14年生．昭和37年北海道大学工学部電気工学科卒業．昭和39年同大学院工学研究科修士課程修了．現在同工学部電子工学科教授．工学博士．計算機応用，ことに日本語文書処理に興味をもつ．電子情報通信学会，日本音響学会各会員．