

結合価に基づく英日機械翻訳の試み

鈴木 雅実 野垣内 出 神 博史
(KDD 研究所)

1 はじめに

トランスファー方式による機械翻訳においては、異なる言語間の接点としての中間表現を、システムの意図に応じて適切に設定することが重要な課題の一つである。このことは改めて指摘する必要もないことかも知れないが、本論で説明する、結合価に基づく言語依存型中間構造を導入する上で有益であると思われるので、幾つかの観点から問題提起を行なってみたい。

1) トランスファーのレベルと中間構造の言語依存性

一般に入力言語の解析のレベルが浅い程、解析結果としての中間構造は表層の言語表現の特徴を強く保持しているためトランスファーの負担が大きいの。また逆に、概念依存構造のような深いレベルの中間構造では、トランスファーの負担は軽くなるが、そこに至るまでの解析に多くの手数がかかるとともに、生成過程において表層上の情報の欠如がネックとなる恐れもある。

2) 中間構造の安定性

述語の表わす意味構造は、統語形式とも密接なつながりがあり、切り離しては考えられない側面も持っている。そこで表層的な表現上のバラエティをより抽象化した統語構造を仮定すれば、その特定の形式の上に意味構造を写像することも可能となる。この抽象化された統語形式に相当するのが、本論で述べる結合価構造である。後で述べるように結合価に基づく中間表現は、意味に近いが統語構造も反映しているため、安定した形式化を可能にする。

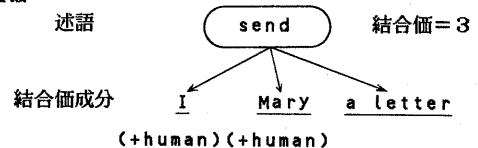
3) 結合価に基づく述語分析の必要性

結合価に基づく中間構造を作り上げるためには、言語固有の述語の特徴を抽出し、それが実現する格形式などについて分析する必要がある。この点については、既存の辞書などでは十分とは言えず、機械翻訳の立場からすれば、なおさらのこと、この観点からの調査・分析が望まれるところである。

2 結合価文法の特徴とその応用

結合価理論(Valenztheorie)または結合価文法(Valency Grammar)については Tesnière 以来の研究の経緯があり、主としてドイツを中心に発展してきた^{2,3)}。要約すれば、結合価文法とは、文の構成要素のうち、述語(動詞および形容詞)に注目し、これと統語的・意味的に結合する、結合価成分(補足語)の数とタイプを記述するための理論である。

例えば



補足語の数については、同じ内容を表わす表現では、異なる言語であっても等しいことも多いが、述語の性質により一致しないこともある。また、補足語のタイプについては主として格形式と統語上のカテゴリーで示されるのが普通である。

表層上の格が明確な言語では、結合価成分は格に対応させる形で定義するのが都合が良い。しかし明示的な格マーカーがやや乏しい英語では述語に対する補足語の形式の設定の仕方に問題がある⁴⁾。またいわゆる補文や不定詞の取り扱いに関しては、表層上の統語的な特徴だけでは次の例のように意味的な結合価を反映することはできない。

- S1) I expect you to do~
 =that you do~
- S2) I promise you to do~
 ≠that you do~

上記のような区別は、英語の場合、Hornbyの動詞パターンなどで示されこともあるが概して不徹底である。⁵⁾

* (上の例では、Hornbyコードは、ともにVP17)

そこで、筆者らは、英語の述語の表わす意味構造をなるべく反映するような、抽象的統語形式としての結合価を、Emons⁶⁾の定式化に従って次のように設定した。¹⁰⁾

S	文
P	述語 [P1(i)(j)で i<j, i, j=2~6、補足成分に対応]
E1	結合価主語
E2	" 直接目的語
E3	" 前置詞句(相当句)
補足成分 E4	" 間接目的語
E5	" 目的格補語
E6	" 主格補語 注)

任意成分 A(A, A, A, ...)

結合価成分の統語形式

NOM	名詞句
IK	不定詞句・動名詞句
ES	補文

先の例文では、次のような関係が成り立つ。

	E1	E2	E4
S1	expect I	you to do~	
	(P12)		
S2	promise I	to do~	you
	(P124)		

個々の述語について、このような分析が必要となるわけであるが、Emonsの理論では、その述語の中核となる意味構造を仮定し、それをもとに結合価を定義するという作業が前提となるので、判断が難しい場合もあり、体系的な分類には膨大な作業を要するものと思われる。

筆者らのグループでは、英語および日本語の基本動詞を中心にそれぞれ約500語づつの分析調査を行ない、結合価パターンと用例を一覧にしたものを活用している。このうち英語動詞については、Hornby動詞パターンとEmonsによる結合価コードとを一部対比させてある。⁷⁾

また一つの文の構造を、述語を中心とする結合価構造を基礎として記述するために、結合価の概念をやや敷衍し、拡張結合価構造(Augmented Valency Structure, 略称 AVS)として表現することにした。

拡張部分は、接続詞を広義の述語として扱っていること、名詞句の内部構造をHead Nounを中心とした依存構造として表わしていること等である。VALANTINEシステムではこのAVSを二言語間の切口として、トランスファーを行なう。

英語のAVS(E-AVS)は、先に説明したEmonsの定式化に従った、述語を中心とする結合価構造を拡張したもので、次の例に示すような表現形式をとっている。

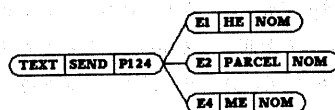


図1 E-AVSの例(He sent a Parcel to me.)

これに対し、日本語のAVS(J-AVS)は、標準的な格フレーム形式を基本としている。統語カテゴリーの違い等を除けば、E-AVSとほぼ同等の構成となっている。

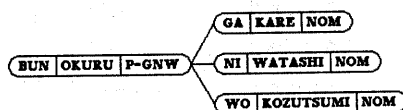


図2 J-AVS(上記E-AVSに対応)

AVSを中間表現としたことによる利点は次の通り。

- 1) 文内容の結合価構造が階層的に表わされており、各ノードには上位および下位ノードとの関係を示す識別子が与えられ、簡潔明瞭な表現形式となっている。
- 2) 結合価構造自体が意味関係の写像でもあり、統語的な特徴もある程度保持しており、多層的な情報を扱える。
- 3) 各ノードはフレーム型の内部構造を持ち、ノード間の情報の授受等により、言わゆるオブジェクト指向の操作が可能である。
- 4) 上記三点の特徴により、結合価成分と述語の組み合わせをもとに、結合価レベルでの見通しの良いトランスファー規則が記述できる。

(注) E1~E6に対応する日本語は必ずしも適切な表現ではないが、ここでは一応の目安として用いている。

3. VALANTINEシステムの処理方式と辞書記述

現在開発中のVALANTINE*システムでは、前章に示したような言語依存型のAVSを中間構造とするトランスファー方式を採用しているが、ここでは、E-AVSの作成およびJ-AVSへのトランスファーを中心に、全体の処理の流れと、辞書の構成についての概略を説明したい。

3.1 システムの構成と処理の手順

図3に全体のシステム構成の略図を示す。ユーザ・インターフェイスの部分については次章で説明する。また、処理の流れは図4のチャートに示す手順に沿っている。

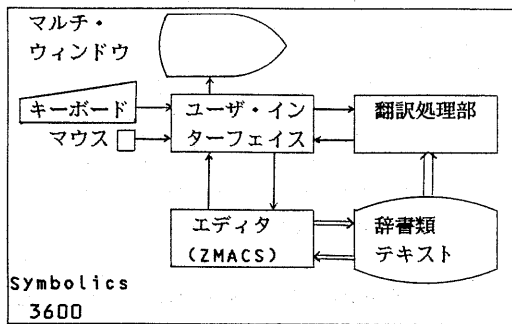


図3 VALANTINE システム構成図

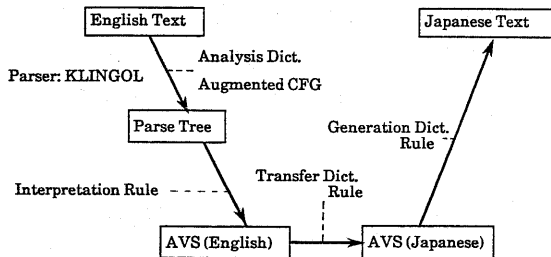


図4 VALANTINE-EJ Process Chart

<解析部>

入力英文は、解析用辞書と文法規則を参照しながら、KLINGOL (拡張LINGOL¹⁾ に手を加えたものをパーサーとして構文解析を行なう。次に、得られた構文解析木に対し結合価解析規則を適用し、英語のAVS (E-AVS) を作成する。

このE-AVSは、前章で示したように、階層的な依存構造となっている。各ノードは上位ノードや下位ノードに対する役割情報の他、自身の統語カテゴリー、属性情報などをもつオブジェクトとして表現されており、その属するクラスのインスタンスとなっている。また、P12、NOMなどのクラスごとに可能な操作がメソッドとして定義されている。

<変換部>

解析処理の結果である、拡張結合価構造E-AVSについて、後に述べるトランスファー辞書内の変換規則を適用して、J-AVSを得る。この処理では、AVSの最上位ノードから、下位ノードへ、順次、depth-firstにトランスファーが行なわれる。変換規則によっては、J-AVS側で、新しいノードが生成されたり、逆にノードの数が減少することもある。

<生成部>

トランスファーによって得られた、日本語結合価構造J-AVSからの生成処理では、やはり最上位ノードから順に下降して depth-firstにノードをたどり、語順規則と接続規則に従って、語が出力される。現在は、アルゴリズムと規則の分離が明確でないこと、Modalityの処理について検討中であることから、現在形能動文のみが出力される。

3.2 辞書規則の構成と内容

VALANTINE-EJで使用される辞書規則類は、図4で示す通りであるが、その特徴を以下に述べる。

1) 解析用辞書(kdict)

- ・記載項目 (見出語を除く)
 - ・品詞
 - ・各種統語カテゴリー
 - Hornby動詞パターン、形容詞の型、共起する前置詞等のマーカー、...
 - ・意味カテゴリー (名詞)
 - +human, +abs, +place 等7種
- ・不規則変化情報
- ・語数 約1,200

* Valency-based LANGUAGE Translator with Interactive Editor

2) 構文解析用文法(kgram)¹²⁾

- KLINGOLでのバースを前提とした書換え規則
- 隣接可能性等の制約条件をadvice部で記述

3) 結合価構造構造解釈規則(interpret)

- ほぼ kgramの規則に対応する 一般的解釈規則と、語に依存する個別規則から構成される。
- 条件部、実行部の記述はリーダーマクロにより、平易な記述を実現

4) トランスファー用辞書(trans-rules)

- 記載項目 (見出語を除く)
- 結合価タイプ (英語)
- 条件部
 - 結合価成分その他の統語的・意味的制約条件
- 実行部
 - J-AVS生成規則[DT]
 - 結合価タイプ (日本語)、結合価パターン
 - 訳語および活用情報
 - 述語の意味的特性[AF]
 - 意味カテゴリー、Modality付加情報
- 語数 約400

例) "send"の辞書記述

```
(send
(P12
(RF +action +ergative)
(DT P-GH (E1 ga) (E2 wo) (okuru 送 R5)))
(P13
(if (+ for c marker [], +human c <E1>, +human c <E3> +
(RF +action +volition)
(DT P-GH (E1 ga) (E3 wo) (yobiniyaru 呼びにや R5))))))
(P123
(if (+ to c marker [], +human c <E1>, +human c <E2> +
(RF +action +causality +volition)
(DT P-GMN (E1 ga) (E2 wo) (E3 ni) (ikaseru 行かせ RU))))))
(P124
(if (+ +human c <E1>, +human c <E4> +
(RF +action +benefactive +ergative +volition)
(DT P-GMW (E1 ga) (E4 ni) (E2 wo) (okuru 送 R5))))))
(P125
(if (+ +human c <E2> +
(RF +action +affective +causality)
(DT P-GWN (E1 ga) (E2 wo) (E5 ni) (ochiraseru 馳らせ RU))))))
)
```

5) 生成規則

主として、語順を決定するための規則、隣接関係から適切な形態素を付加するためのもの、Modalityを生成するための規則に分かれるが、先にも述べたように、アルゴリズムとの分離等について検討中である。

3.3 処理例

以上の処理方式および辞書記述をふまえた上で、実際の例文に即して、翻訳処理の説明を行なう。

S3 THE ACTUAL TRANSMISSION CONSISTS
IN SENDING A FREQUENCY PAIR.

S3の例文に対してまず、構文解析を行ない、次の構文解析木を得る。ここで主動詞consistは、Hornby動詞パターンVP3A、その前置詞マーカが" in"であるという情報をもつ・

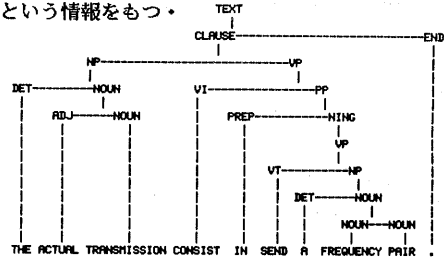


図5 S3のParse Tree

次にこの解析木に対し、結合価構造解釈規則を適用する。即ち、最初のNPの部分がE1に、PPの部分がE3に対応し、consistはこれらの2つの結合価成文を支配する述語P13として解釈される。(E3の前置詞マーカは" in") この場合、E3自体も、動詞" send" を述語とする結合価構造を含んでいるので、その部分が展開され、結果として次のようなE-AVSが得られる。



図6 S3に対応するE-AVS

E-AVSは、英語側の中間表現であり、述語を中心とした依存構造となっているので、最上位の述語をkeyとしてトランスファー辞書を引き、この場合のP13となる結合価タイプを捜し、その内容を実行する。

consistの辞書記述は次のようになっている。

```
(consist
  (P13
    (if (← of_e_marker[] +
          (RF +state +relation)
          (DT P-GK (E1 ga) (E3 kara) (naru な R5)))
        (← in_e_marker[] +
          (RF +state)
          (DT P-GN (E1 ga) (E3 ni) (aru あ R5)))
      -----
    )
  )
)
```

この場合は、E3のマーカ―が” in” であるので、上
の下線部が適用され、翻訳規則に従って変換が行なわれ
る。同様の操作が、E3の中の” send” についても実行
される（sendのトランスファー規則については前節を
参照のこと）。述語の結合価成分についても、順次トラ
ンスファー辞書の中の規則が展開され、E-AVSは最終
的に次に示すようなJ-AVSに変換される。

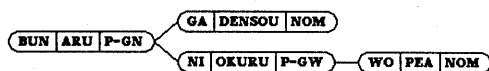


図7 s3に対応するJ-AVS

J-AVSからの生成処理では、日本語の格形式による
語順に従って語を出力しているが、ここでは「ある」に
対する、に格に相当するのが「送る」という述語なの
で、名詞化するための「こと」が付加されている。

実際の伝送は周波数ペアを 送ることにある。

— 翻訳例 —

4. VALANTINEのユーザ・インターフェイス

機械翻訳システムの研究開発途上では、特に言語デー
タ（辞書規則類）の蓄積および妥当性のチェックに相当
なマンパワーを必要とし、効率向上のための支援環境を
充実させる必要性が指摘され始めている。^{13, 14, 15)}

また、実際にシステムを運用するユーザを想定したマ
ン・マシン・インターフェイスの設計も、当然プロトタ
イピングの時点から考慮するべきであろう。¹⁶⁾

VALANTINEシステムでは、以上の二点について、現
在のところ、次のような使用状況を想定して、ユーザ・
インターフェイスの設計を行なっている。

1) 小量文書の逐次翻訳

内容は特に技術文書に限定しない予定であるが、

大量文書の一括翻訳ではなく、翻訳要求が時間的に五
月雨式に生起する状況。

2) 翻訳の専門家による使用

翻訳の素人ではなくて、ある程度2言語の知識に
通じた人の使用を想定。後編集も翻訳者自身が行な
う。

3) 翻訳者の操作可能範囲の拡大

会話式のエディタ機能により、辞書内容の修正・
追加を容易にする。

4) システム開発環境の維持

翻訳システム的能力・機能向上のため、必要に応
じた中間結果の表示、トレース機能等を備えておく。

5) 通信・ファイル管理機構の強化

将来、オンラインでの翻訳対象文書および翻訳結
果の通信等の需要が多くなることが予想されるので、
これに対処するための通信機能、言語データ等のファ
イル管理機能を充実させる。

この他にも考えるべき要因は多いと思われるが、現段
階では上記項目の3)と4)を重視してユーザ・イン
ターフェイスの試作を行なっているので、その内容を紹
介したい。

4. 1 マルチウィンドウ画面構成とメニュー選択

VALANTINE-EJは現在、システムの処理アルゴリズ
ム自体を改良しつつあるため、デバッグ情報は不可欠で
ある。また、言語データの妥当性のチェックのためにも、
中間結果を分析する必要がある。このため、入力
文・出力文の表示の他に、構文解析木・AVSの表示およ
び、その他のデバッグ情報の表示を、図8で示すような
マルチウィンドウによって構成している。複数の情報を
同時に見ることができる点が有利である。

また、通常のコマンドの実行の殆んどは、マウスでの
メニュー選択によって行なえるようになっている。（図
8参照）

図8 VALANTINE-EJの画面構成と実行例

VALANTINE - EJ 1st Special Lab. KDD
Multi Language Systems

Parsing time = 266 ns, Response time = 1 sec.
 Sen time = 0 ns, Response time = 0 sec.
 Describing # <NOM 42750571>
 Governor : TRANSMISSION
 Superior : # <P13 42750737>
 Arbitraries
 PQ : (# <DET 42750523> # <P1 42750546>)
 RQ : NIL
 Feature :
 Message :
 CASE : NOMINATIVE
 SYNTAX_NODE : NP

Select Pane
 English Tree Structure
 Text Input Pane
 Eval Input Pane

<画面選択メニュー>
 <この選択により、AVS表示の代りに
 Parse Treeが表示される>

<入力文>

Input Text
実際の伝送は周波数ペアを送ることにある。

<出力文>

Output Text

English Valency Structure

< E-AVS表示 >

Japanese Valency Structure

< J-AVS表示 >

4.2 AVSのグラフィック表示

E-AVSおよびJ-AVSは、前章までに説明したようなグラフィック表示で、それぞれのウィンドウに現われる。通常は、述語ノードの直下のノードまでの範囲が表示される。各ノードは Mouse Sensitive (マウスのポインタを近づけると、各種の操作ボタンにより可能) となっており、まだ表示されていない下位ノードを表示(open)したり(下図参照)、ノード自体が持っている各種の属性情報等をリストアップする機能が付加されている。(図8左上informationのウィンドウ)

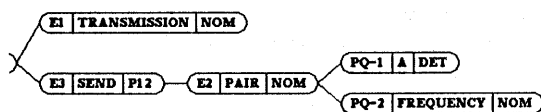


図9 "pair" (図6参照)のopen表示

4.3 実行モードの変更とトレース機能

メニューからの選択により、解析木表示をしない実行モードへの変更が可能である。(現在AVSは必ず表示) また現在、構文解析については、文法規則適用経過のトレースのON/OFFが選択できる。またその際、任意の統語カテゴリー(複数でも可)のみの条件付トレースも可能である。

4.4 辞書規則類のup-date機能

翻訳処理仮定の子エックによっていずれかの辞書規則ファイルに、追加・修正を施したいという要求が生じた場合、トップレベルのメニューで"Update Files"を選択すると、次のメニューに、言語データファイルの一覧が現われる、この中から更新すべきファイルをマウスで選択し、"Edit"を指定すると自動的にシステムのテキスト・エディタ(ZMACS)に該当ファイルが引き込まれるようになっている。必要な修正を行なってファイルをsaveした後、再びマウスで"Load"を指定すると、システムの参照ファイルが更新後のものになる。このエディタについては、よりインテリジェントな機能を付加する方向で開発中である。

5. 検討課題と考察

現在のVALANTINEシステムでの問題点、および今後インプリメント予定であるが検討中の項目について考察を行なってみよう。

5.1 動詞の分類一意味的特徴と統語的特徴の関連一

2章の例でも示したように、述語動詞の意味的特徴を統語構造の上に反映させようとすると、若干の無理が生じ、このためEmonsの定式化によって、6種の補足成分のカテゴリーを設定した。しかし、この分類でも、なおあいまいさは残り、特にE2については、受け身可能性等の統語的特性だけをとってみてもさらに下位区分の可能性もある。実際Allerton⁸⁾は、目的語をobjectとobjoid(受動化できない疑似目的語)に区別する等、意味的役割-結合価構造-表層構造の多層分析に基づく動詞辞書を構想している。VALANTINEシステムでは、トランスファー規則の細分はなるべく避けるという方針で臨んでいるが、場合によっては結合価成分のカテゴリー数の増加も有り得る。

5.2 格文法の利用と任意成分の取り扱い

VALANTINEシステムでは、述語の補足成分については、結合価のレベルでトランスファーが行なわれるため、各補足成分に深層格ラベルを割り当てるような処理は行なっていない。しかし、前置詞句を中心とする、述語の任意成分については、多義性等の問題から、その意味的役割(即ち格ラベル)をある程度決定しなければ、適切な訳語を選択できないという問題がある。¹⁷⁾

しかるに、格ラベルの決定には多くの場合、その任意成分を支配する述語の意味的特徴および任意成分自身のgoverner wordの意味属性等の情報を参加させる必要がある。従って、これらの属性なり意味特徴を述語や名詞句に付加する方法が問題となる。例えば、述語の意味的特徴としては、1) 補足成分の数とタイプに関わるもの、^{9, 1)} 2) 共起性の高い任意成分に作用するもの、¹⁸⁾ 3) Modality成分の形式・意味に関係するもの¹⁹⁾等が考えられる。名詞の意味属性についても様々な観点からの分類が可能と思われるが、利用しているのは一部に過ぎない。

前の1)については、結合価コード自体がそれに該当するわけであるが、トランスファー辞書のAF(Analyzed Feature)の項目には、その場合の述語の意味特徴を、属性名のリストとして記入してある。これらは、必要に応じて任意成分の役割決定やModality成分の処理に役立てる予定であるが、補足成分の格ラベル決定と生成処理への反映に使うことも意図している。²⁰⁾

5.3 Modality成分の処理

前節で述べたように、英語のModality成分の処理と、日本語側での生成にも、述語の意味特徴の関係する側面を抽出する必要がある。特に難しいのは、英語の場合助動詞の多義性等であるが、文脈の影響を考慮しなくてもある程度蓋然性の高い処理規則を書くべく検討中である。またトランスファーについては、英語と日本語のModalityの相違や、述語自体に含まれるModality的な意味特徴(例えば、能力・義務etc)の関与等について調査を進める予定である。さらに、独立分詞構文など、従来の格文法では捉えきれない構文要素についても、広義のModalityとしての取り扱いの可能性を考慮すべきであろう。

6. まとめ

言語依存型の間接表現として結合価構造を用いた、英日機械翻訳の試みについて、その概要を述べた。統語構造と意味構造の間において、多層的な情報を吸収する結合価構造の有効性を示したが、その特徴をさらに活かすため、システムの改良と言語データの蓄積を図る予定である。

謝辞

結合価理論に関する興味深い資料を提供して下さった、備前IRの若菜氏、井倉氏他の調査グループの方々、ソフトウェア作成上でたいへんお世話になった、備前SCCの田中氏と向平氏に厚くお礼申し上げる。

また、システム設計等に対し貴重なコメントを頂いた、グルノーブル大学のViquois教授、シュトゥットガルト大学の羽中田博士に深く感謝する。

最後に、日頃御指導頂くKDD研究所鍛冶所長、野坂副所長、榎松次長、武田第一特別研究室長に感謝する。

参考文献

<結合価・言語論関係>

- (1) 村木 新次郎 他(1984) 辞書における格情報の記述、自然言語処理研究会46-3
 - (2) 石綿 敏雄 他(1983) 結合価から見た日本文法、朝倉日本語新講座3(文法と意味I)
 - (3) H.Weinrich, 脇阪 豊 他訳(1984) 言語とテキスト(3章)、紀伊国屋書店
 - (4) 小西 友七 編(1980) 英語基本動詞辞典、研究社
 - (5) A.S.Hornby, 伊藤健三 訳注(1977) 英語の型と語法、オックスフォード大学出版局
 - (6) R.Emons(1974) Valenzen englischer Prädikatsverben, Niemeyer, Tübingen
 - (7) 第一特別研究室(1984~1985) 英語動詞結合価調査表 他、KDD研究所内部資料
 - (8) D.J.Allerton (1982) Valency and the English Verb, Academic Press
 - (9) J.Rickmeyer(1977) Kleines Japanisches Valenzlexikon, Buske, Hamburg
- #### <システム・開発環境>
- (10) 鈴木 雅実 他(1984) 結合価に基づく言語間トランスファーのための一考察、情報処理第29回全国大会4N-1
 - (11) 田中 穂積(1979) 計算機による自然言語の意味処理に関する研究、電総研研究報告797号
 - (12) 橋本和夫 他(1983) 英語構文解析のための文法作成、電子通信学会総合全国大会1331
 - (13) Ch.Boitet, et al.(1984) Expert Systems and Other New Techniques in MT Systems, COLING'84, 468-471
 - (14) 中村 順一 他(1984) マルチウィンドを利用した機械翻訳のための文法開発ツール、情報処理第29回全国大会4N-2
 - (15) 小暮 潔 他(1984) 辞書編集用フレームエディタ、自然言語処理研究会45-1
 - (16) A.K.Melby(1984) Recipe for a translator work station, Multilingua 3-4
 - (17) 高松 忍、西田 富士夫(1981) 動詞パターンと格構造に基づく英日機械翻訳、信学論Vol.J-64-D, No9, 815-822
 - (18) 辻井 潤一 他(1985) 日英翻訳過程での処理とその翻訳結果への反映、自然言語処理研究会47-10
 - (19) 野垣内 出 他(1985) 動詞意味素性の付加による日本語文アスペクトの解析および翻訳について、自然言語処理研究会48-4
 - (20) 鈴木 雅実 他(1984) 英文の深層格推定のための動詞分析、情報処理第28回全国大会6L-4