自然言語処理 49－4
（1985. 5. 24）

語彙遷移ネットワーク文法について
機械翻訳システムTAURASの意味分析方式

天野真家　野上宏康　三池誠司
（株）東芝　総合研究所

1．はじめに

自然言語処理 … わけても、統語－意味分析 … の困難さは自然言語が非常に多くの情報を担っていることに、あるいは人間知性の表象であることに由来していると考えられるが、しかし、そのような、いわば質的な複雑さのみに由来している訳ではない。質的にどのように困難であろうとも、その扱う領域が小さく制限されていれば、それを扱う方法はあるものであることはSchankやCharniakあるいはTAUM-METEOの成果を見ればあきらかであろう。例えば、次のような文の翻訳を考えてみよう；

(a) She was Catholic. Not a Catholic, but Catholic.
(b) 彼女はカトリックだった。ただのカトリックではない。筋金入りなのだ。

…PAUL ERDMAN 著　池　央耿　訳　The crash of '79 より

(a) の文を(b) のように翻訳することは、一般的には現在のレベルの機械翻訳システムでは不可能と言って良いであろう。それにもかかわらず、この文に限れば、"Catholic"が無冠詞であるか、冠詞"a" を持っているかをきっかけにして(b) のように翻訳することは現在のレベルでも可能である。換言すれば、個々の言語現象はそれが、個々に挙げられる限り解く方法があるということである。一般的に解けない理由は、それらの言語現象が一挙にあげられる（それらを扱う文法あるいは意味規則間の相互干渉が複雑に絡みだす；相互干渉性）と共に、全てが挙げ尽されていない（計算言語学の見地から全ての言語現象が予め見えている訳ではない；不透明性）からである。

自然言語処理を実用の域にまで持ち上げるのを阻んでいる大きな壁は質の問題というより、むしろ相互干渉性と不透明性という２つの性質を生ぜしめている量の問題である。より正確には、量を裏に秘めた質の問題であろう。量に具現を阻まれないrobust algorithmを見出すことこそ、自然言語処理における最も本質的な課題であろう。

2．動機

自然言語処理の大目標のひとつに機械翻訳がある。周知の如く、翻訳という作業は極めて知的な作業であって、単なる語の置換ではない。従ってSYNTAX－ORIENTEDな翻訳というものはあり得ないであろうということは予想に難くない。勿論、極めて近い親族関係にある言語対で、構文的曖昧性に関しても、語彙的曖昧性に関しても同一性質を持ち、それらを解消する必要がほとんどない程なら、

実用上、SYNTAX-ORIENTED な翻訳システムも可能であろうが、一般的ではない。意味分析は、一方、非常にコストがかかる。従って、意味分析の再試行を繰り返すような方式にはしたくないという要求がある。語彙遷移ネットワーク文法は計算の複雑さが冪で利く部分は単純な処理にし、複雑な意味処理は計算の複雑さが線形の部分に置くことによって全体のコスト・ダウンを目ざすものである。

3．　モデルのありかた

用語の定義：

構文分析；　語列を受理又は拒否し、受理した語列に関しては、予め定められた構造を構築すること。統語的であるか、統語－意味的であるかには言及しない。

統語分析；　比較的少数の語のクラスと対象言語の統語的特性だけを用いておこなう構文分析である。勿論、語のクラスをいわゆる意味文法的に設定した場合はどうなるかということになるが、厳密に統語論と意味論の境界線を引くことができないので、この定義もその程度の曖昧性を持つ。

統語－意味分析；　統語分析と意味分析の両方を含んだもの。どのように、この両方の分析を行うかには言及しない。

パーザ；　構文分析を行う機構。

ここで述べる試論は統語－意味分析の原理的な能力の問題に関したものではない。原理的な能力の問題としてなら、自然言語の文が帰納的可算集合をなすと仮定すれば、最近のパーサは大抵の場合、０型となっており、それを受理する原理的な能力はあるのである。しかし、機械翻訳のような実用システムを目指している時、そのような（パーザの）存在証明は何の役にも立たない。この場合、“実用的”という事は“原理的”の彼岸にあるとさえ言えるだろう。また、存在証明のみならず、実際に、そのようなシステムを机上で構成したとしても、それだけではなお不十分である。実用的なシステムには種々の制限が課せられる。パーザを搭載する計算機の記憶容量、処理速度、文法記述の容易さなどが、それである。この種のモデルの正当性は、従って原理的証明だけでは不十分であり、実システムの構成によって証明されるのみである。世に種々の文法理論があるが、工学的手法に…　心理学的側面に興味を持つなどでなく　…重点を置きながら、なお工学的実現可能性を持たない理論には理論としての価値はない。Schank、Charniak 等のモデルは、複雑な個々の言語現象を説明するモデルにはなっているであろうが、自然言語処理の問題は既に述べてきた如く、それらの問題が個々に解決される方法ではなく、一挙に生じた時に対処できる…換言すれば大規模データに対処できるモデルでなくてはならない。小規模モデルの単なる拡張で大規模モデルが自動的に構成できる保証は何もないのである。

工学的実現可能性について

以下に示す［意味は語彙的である］という仮説は、意味は統語規則に書くべきでないという禁止条項として述べてある。その理由は一般的な縮退による困難性として挙げられている。従って個々の方法論を一々検証する必要はないものである。しかし、ここに述べることは原理的可能性の問題ではなく、秀れて工学的実現可能性の問題であるから、統語規則に意味分析規則を付加する方法も“原理的に可能”ではある。この様な問題を取り扱う場合、原理的に可能ではあっても工学的実現可能性がないという概念が必要であろう。工学的実現可能性の証明要件は単なる原理的可能性証明あるいはalgorithm の存在証明だけではなく、先に述べた２つの要件…大規模データの相互干渉性と不透明性…の解決に関するパフォーマンス、コスト、メンテナンス、などが関係しよう。

４．　TAURASの意味モデル

統語－意味分析システムを構築しようとした場合、我々が用いる数万の語そのものに関する統語規則を作ろうとすることは事実上不可能であろう。従って、語をクラスに分け、そのクラス間の関係を規則化することは自然な考え方である。しかし、語をそのようなクラスに分け、語そのものの持つ種々の属性を放棄し、クラスとして扱うに至った瞬間から、自然言語の文は統語上の曖昧性を抱えこむことになる。関係というものは、本来、線形に並べて曖昧性無く表すことはできない。その曖昧性を解く鍵となっているのが、各語が内包している概念であり、我々は、文で述べられている関係を現実の世界へ写像することによってその曖昧性を除去している。しかし、語をクラスとして抽象し、かつ、語が表す概念の関係を線形に並べるという２重の縮退を施すことは必然的に解決できない曖昧性を産み出す。線形の文を２次元的構造に復元するには、概念関係の助けを必要とすることは当然である。

仮説１；　意味は、記号の集合から、概念の集合への写像である。

［従来モデル］

意味は、記号集合中の対象から概念集合中の対象への写像として定義する。統語－意味分析においては記号はクラス名（品詞といっても良いが、伝統的文法でいう品詞に限定されるわけではないので、より一般的にクラスということにする）として現れる。概念集合は現実の世界そのもののモデルであってもよいが、それに制限される必要はない。現実世界を当てた場合、不思議の国のアリスの世界のような想像上の世界をどのように処理するかの問題が生じるので、ここでは、概念集合は一種の公理系として存在する。アリスの世界、現実の世界はプラグマティックスとして、多世界問題として考えればよいと思うが、この問題に関しては将来の問題として、ここでは触れないでおく。従来の意味標識と選択制限規則による意味処理は統語分析中に局所的にこの写像を行うことにより構文の曖昧性を除去しようとするものであった。

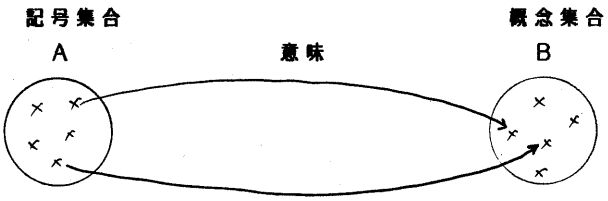


図１　意味の従来モデル

この記号集合Ａにおける文の表現は曖昧性があっても概念集合Ｂへの写像（意味）によって簡単に除去できる。

例；

I saw a girl with a handbag.　[with以下は　girlにかかる]

しかし、語のままでは法則化が繁雑で事実上できないので、抽象化してクラスの集合Ａ´を作る。この瞬間、文の曖昧性を解くべき概念集合への写像の道が断たれる。

例；

pn vt det noun prep det noun.　[prep　以下の係り先が不明]

この写像の道を回復するため、意味標識のような局所的な方法によるのではなく、Ａ´から一旦、Ａに戻り、それからＢへの写像を施す。

仮説２；　意味は基本的に語彙的である。[TAURASの立場]

既に述べて来たように、意味は記号集合から、概念関係の集合への写像である。一方、文法規則を書く場合、既に述べたように語を直接、要素（終端記号）として文法を書くことはできないので、クラスに分類し、このクラスで文法を記述することになる。文法記述の要素であるクラスの数はすくない方が、文法が簡潔になってよい。これは必然的に大量の概念の一つの記号への縮退を惹き起す。名詞というクラスを例にとれば、数万の概念が一つのクラスに縮退している。これが仮説１のモデルによる統語－意味分析システムの構築を困難にしている。この縮退による困難を避けるため用いらる方法は、一つは品詞細分である。これは本質的に解にならないことはあきらかだろう。他の効果的な方法は条件文を文法規則に付加することである、しかし数万の語がそこに縮退している規則に条件文を付けて再びクラスから語そのものと々と転化させることは、必然的に条件の増加をもたらし、統語規則そのものよりもはるかに巨大な条件文の羅列となる。即ち、表面上統語規則でありながら実は語による意味規則となってしまう。これが意味は基本的に語彙的であるという意味の一である。

上述したように、統語規則を意味規則の海の中に沈めてしまうことは文法記述の容易さを害するものである。意味は語彙的であるとは、意味規則は統語規則の側ではなく、語彙の側に書くべきものであるという意

味でもある。

仮説３；　意味は統語構造の集合と語彙的意味の集合の直積から概念構造の集合への写像である。［TAURASの意味モデル］

ここで扱う意味分析は、意味標識と選択制限規則によるような局所的なものでなく、総合的に行うことを目指している。従って、仮説１の記号集合を少しくmodifyして記号構造の集合を作る。これは統語論　－　記号と記号との関係の法則　－　により容易に構築可能である。統語論により記号間の関係は完全に記述され統語構造の集合ができる。この世界はSYNTAX-ORIENTED なモデルによる構文分析ならばこれが直接に構文分析結果となる。しかしながら、既に述べたように２次元の関係を１次元に表したことにより生ずる曖昧性はここではまだ除去されていない。その曖昧性を解く鍵は概念集合にあり、統語構造の集合と語彙の集合の直積から、概念集合への写像によって除去されるというのが本仮説である；

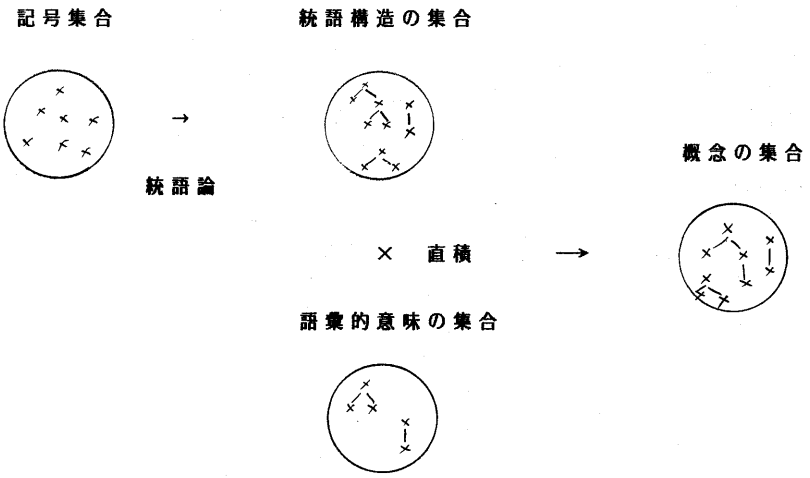


図２　TAURASの意味モデル

この場合、概念集合への写像には、勿論、統語構造が関わっているが、より本質的に語彙が関わっていることを主張している。

より本質的に語彙が関わっている例として、統語構造だけでは、意味的曖昧性の除去ができない例は容易に作ることができる。

例；

自動車はハンドルは持っている。

ハンドルは自動車は持っている。

この例でも分かるように、上記のPART-OF の関係は統語構造には何の関係もない。

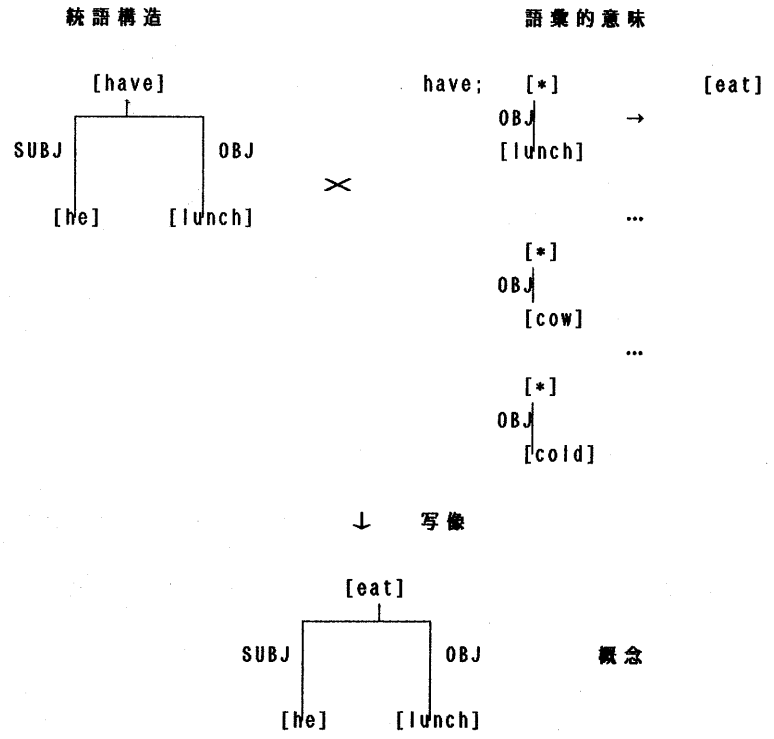
5． 意味の事例とその分析

本節では種々の意味の形態のうち、典型的なものに関して、その例を示す。

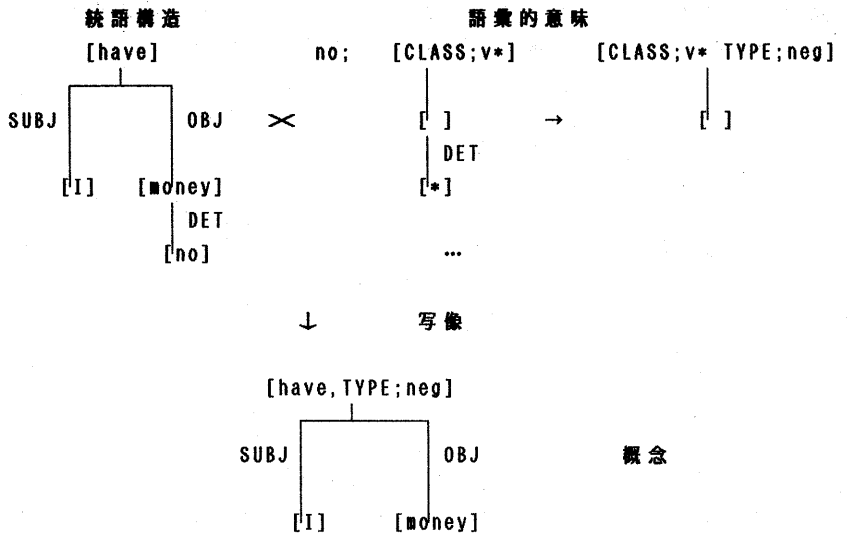
記号法；統語構造の各終端記号は[--]で表す。これらは辞書項目及び、形態素分析、統語分析の結果判明した属性の束であるが、ここでは説明に必要なものだけを表記している。ここに示した表記は応用の概念を得るためのもので、厳密なものではない。例えば、"have"の意味として"eat" としているが、これは本来英語の"eat" ではないが、その主要な意味で代用している。

*; self-reference
CLASS; 語のクラスを示す
TYPE; 文のタイプを示す
SUBJ, OBJ 等のアーク名； 統語－意味的関係を示す
v*の*; wild card character
[ ]; 任意要素

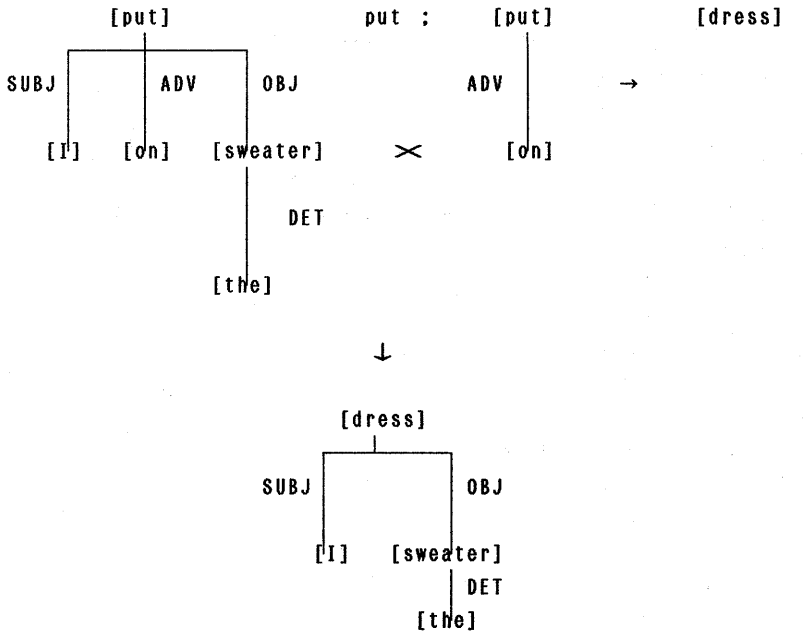
語の意味；



構造に関わる意味；



慣用句の意味；



６．インプリメンテーション

これまでに示した意味モデルによる統語－意味分析システムは語彙遷移ネットワーク文法として実現されている。語彙遷移ネットワーク文法は一般文法部と語彙文法部の２部門からなる。その構成は図３のようになっている。

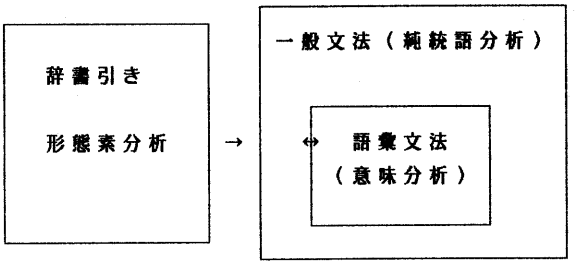


図３　語彙遷移ネットワーク文法の構成

一般文法；　一般文法部は表記形式としては拡張遷移ネットワークを改良したものを用いている。但し、文法記述の形態は先に述べた様に統語的である。これを特に従来の意味処理が非体系的、部分的に入った構文分析と区別したい時には純統語分析と呼ぶことにする。ここでは、語彙の意味による差異をその統語構造には反映させない。従って、著名な例として、例えば、

a)　I persuaded her to go.
b)　I promised her to go.

の２文は全く同一の過程を通って分析され、同一の統語構造を持つ。即ち、これらは語のクラスの並びとして見た場合。

c)　pn vt pn to vi.

となり、統語的には何ら差異がないからである。同様にして次のa), b)はそれぞれ、c)として同一の構造を持つ。

a)　He is able to speak French.
b)　I am glad to see Jack.
c)　pn be adj to vt noun

a)　I want to go there.
b)　You have to come here.
c)　pn vt to vi adv

このようにして、計算の複雑さが幂できく統語分析を簡単化している。

語彙文法：

語彙文法は各語に付されている意味規則である。その主たる役割は；

① 一般文法で決定できない語の意味を決定する
② 一般文法で解消されない構文的曖昧性を解消する
③ 一般文法で判定できない意味の真偽性を判定する

の３つである。これらは全て次の形式の規則によって統一的に行われる。

1) ＭＰ→ＴＰ；［コントロール］（［コンディション］［アクション］）

2) ＭＰ；［コントロール］（［コンディション］）

（）内は省略が許されることを意味する。

ＭＰ； マッチング・パターン。ネットワークで表現される。一般文法部の出力結果に部分照合される。

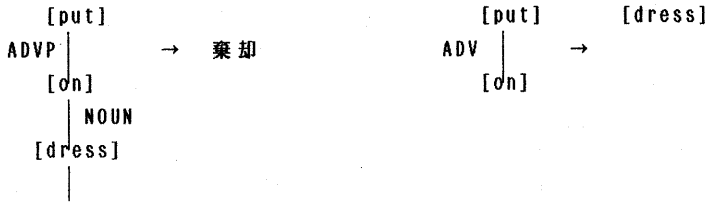
ＴＰ； ターゲット・パターン。ネットワークで表現される。ＭＰと照合がとれた一般文法部の出力結果はこのパターンに変換される。

コントロール； 規則の適用の制御である。上記1)では規則適用を一度で終了するか、全て試みるかの２モードがある。2)ではＭＰに照合する部分パターンを持つネットワークを受理するか、棄却するかの２モードがある。

図３において、一般文法部と語彙文法部の間が→でなく、↔で結ばれているが、これは、ある解釈が意味的に不整合で棄却された場合、一般文法部が他の解釈を再び語彙文法に実行させるためのものである。これは多品詞による統語構造の曖昧性を除去するために上記2)の機能によって実行される。例えば、

I put on the dress.

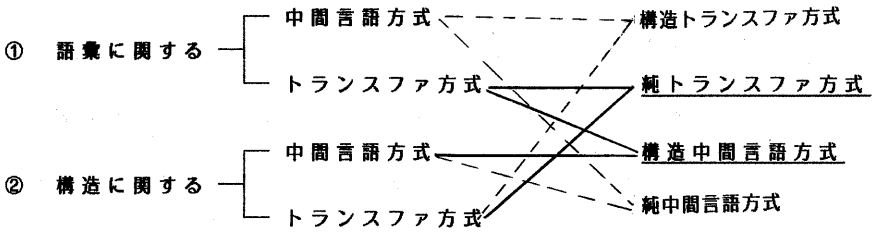
の"put" に注目すると、自動詞、他動詞の２用法がある。自動詞の場合、"on"は前置詞で"on the dress"は副詞句となって"put" を修飾する構造が得られ、他動詞の場合、"on"は副詞となる構造が得られるが自動詞の用法は棄却される；



1) 自動詞の用法　　2) 他動詞の用法

７．　応用　…　機械翻訳

語彙遷移ネットワーク文法は機械翻訳システムTAURASに用いられている。筆者等はTAURASを便宜上トランスファ方式と呼んでいるが、その最初の文献1)以来、これはヨーロッパで意図されたトランスファ方式ではないことを主張している。トランスファ方式が有効なのは、構文的にも語彙的にも類似関係にある同系の言語対に対してであって、日本語とヨーロッパ諸語の様にその形態が隔絶している言語対ではその本来の意味での有効性は発揮されない。即ち、効率の良い高品質翻訳は不可能である。従来、翻訳方式として、トランスファ方式と中間言語(PIVOT)方式に分類されていたが（電総研の融合方式は命名の観点が異なっており、同じ観点に立てば、上記２つのいずれかあるいは、以下に述べる方式になる）もう少し細分した方が良いと思われる。即ち、



従って、４種の可能性が生ずるが、語彙に関する中間言語方式をとっている実用翻訳システムはないようである。TAURASは構造中間言語方式を採用している。日本語と世界諸国語間の多言語翻訳を行う時これが、最も効率の良い方式と思われるからである。TAURASでは語彙文法によって語の意味が決定される。この時、決定される語の意味を中間言語の語彙で表現すれば、語彙に関する中間言語方式となる。語彙は通常膨大な数になるため、中間言語の設定が困難であると共に、原言語が１つで対象言語が複数の時は、　原言語　→　中間言語　→　対象言語　という繁雑なルートより、　原言語　→　対象言語　のほうが効率が良い。この後者の方式は語の意味を決定する時、中間言語に変換せずに直接、対象言語に変換していることになり、その意味でトランスファ方式である。一方、文の表す構造は、語間の関係が概念関係になっている。関係の概念は、語彙に比べれば少数ですみ、中間言語を設定しやすい。また、純統語関係では文の意味が表せえない事は明白である。この意味でTAURASは構造中間言語方式になっているのである。

８．　おわりに

TAURASの記述に際して、種々の概念を必要とした。工学的実現可能性、純統語分析、純トランスファ方式、構文中間言語方式などである。これらが混沌としている機械翻訳の分野を多少とも整理できれば幸いである。

参考文献；1)　天野、平川；　英日機械翻訳用パーサについて、情処、NL研技術資料 32 , 1982
2)　天野、野上et al;　LEXICAL NETWORK GRAMMAR,信学会、総合全国大会,'84
3)　野上、熊野et al;　TAURASの構文解析について、情処第30回全国大会、'85
4)　三池、野上et al;　TAURASの語彙文法について、情処第30回全国大会、'85