

国立国語研究所学術情報リポジトリ

An automatic syntactic system of natural language

メタデータ	言語: jpn 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 石綿, 敏雄, ISHIWATA, Toshio メールアドレス: 所属:
URL	https://doi.org/10.15084/00001047

変形とその逆探知を含む構文解析

石 綿 敏 雄

0. 概 要

自動構文解析は各種の言語情報処理の基礎となる，ということができよう。たとえば内容分析 (Content analysis)，質問解答 (Question answering)，機械翻訳などを行うためには，必要不可欠であるといえる。そのような実用を目的としなくても，言語の構造あるいは言語行動の構造を実験的に探索するというような目的に対しても，有効な方法であるといえよう。そこで，さきに考えた構文解析を含むシステム AUTOSEG のあとを受けて，あらたに構文解析のプログラム (IRIS0A と命名) をつくり，いくつかの実験を試みた。ここではそれについて報告する。AUTOSEG との主な相違点をあげれば，AUTOSEG では扱う文法の型が context free の句構造規則の取扱が可能であるにすぎなかったのに対して，IRIS0A ではそれを含んで変形規則までを扱うことができ，AUTOSEG では取扱言語を主として日本語に限定していたのに対して，IRIS0A は日本語を含めて他の型の言語も同時に扱うことを想定して設計したことなどである。

IRIS0A でも AUTOSEG と同様，辞書・文法とプログラムとは一応独立とした。語彙・文法の記述者が，自己の信ずる理論に従って文法と語彙の記述を行うと，プログラムはそれをいわばパラメータとして，その言語記述者が記述した範囲内で与えられた文の解析を行うのである。AUTOSEG では，いわゆる句構造の範囲内で記述の自由を許している。従って橋本文法の連文節や時枝文法の入子型構造は処理できるが，変形の操作まではできない。それに対して，IRIS0A はいわゆる伝統文法の範囲 (かかり受けだけを扱う) を始めとして，

句構造文法の規則あるいは結合価文法はもちろんのこと、変形規則の扱いも含むので、かなり多くの型の文法の取り扱いが可能である。そこで、このような諸種の型の文法による言語記述をパラメータとしてプログラムを実行すれば、どのような型の文法による言語記述がどの点でどのような働きを示せるかを、具体的に明示することができる。実験の結果では伝統的な国文法では、ことばのかかり受けを重視し、その点の解明ははっきりと行うことができる。これに対してたとえば結合価文法では表層構造に現われない要素で、深層ではあったと考えられるものを、ある程度機械的に示すことができる。この点は伝統的な国文法ではできない。このようなことが示せるのは、プログラムと、解析に用いる言語記述（辞書と文法）をはっきりと分けたからである。

そのことはまた、日本語という特殊な一言語に処理を限定しないようにすることを可能にした。この点からいうと AUTOSEG もすでに句構造の範囲で、一般の言語に適用が可能であったはずであるが、そのことを十分意識していなかったために、必ずしも十分とはいえなかった点がある。IRIS 0 A の設計にあたってははじめからそのことを考慮に入れてあったので、プログラム作成者の意図した範囲では一応の目的を果たした。このプログラムではたとえば埋めこみの文の存在とその原形を推測するためには表層の形を変えてみるということを行うのであるが、その操作も「変形」のなかに含めている。その際、あるいは他の場合に、表層の形がどのような句構造規則にどのような変形規則をあてはめて実行したのか（変形規則の実行は何回か重ねて行われる可能性もある）を推測ないしは逆探知をするテーブルを作成し、これを順次に行わせるので、ある特定の言語のために特定の、いわば ad hoc なルーチンを用意するということはしない。したがって言語記述の部分は完全にプログラムと独立している。

このようにしたために、言語記述の部分が特定言語の特性の記述になると共に、プログラムの方は言語行動のシミュレーションの方へ発展させる第一歩と考えてもよいことになった。ただしきわめて幼稚で初歩的な第一歩であって、これから真の姿を想定するには、まだ相当の距離がある。それに対して言語記

述の方は、このような段階でもある程度利用可能であろう。日本語についての筆者自身の整理は、国語研報告54所収の「日本語の生成語彙論的記述と言語処理への応用」(文献1)およびそれを準備するためのいくつかの稿を利用した。このプログラムの設計に当たってもその点を基礎として考えている。プログラムの実験には、日本語の場合は変形操作を伴う結合価文法 (Valenztheorie) を主としていくつかの文法理論による記述を試みたが、ドイツ語、フランス語を扱うときは、変形操作を伴う結合価文法によって記述し、その結果を比較してみた。その結果は、日本語では名詞などの inherent features と selectional rule がきわめて重要な働きを示すのに対し、ドイツ語やフランス語ではそれ以前に性数格などの文法範疇が構文的にきわめて強力に働き、時には inherent features をわざわざ持ち出さなくとも解析の手續をすましてしまうケースが少なからずみられた。これはやや高度な見地からの言語構造の比較であるが、その前にも、動詞句の構造の特性などは同じフォーマットの記述の上に歴然と現われる。たとえば日本語の動詞句では支配語である動詞の左側にすべての要素が来、それもまずコネクタとしての助詞が来、次に実詞としての名詞がくる。フランス語ではこれらは動詞の左側にくるもの(主格)と右側にくるものに分かれる。名詞句の構造は、ドイツ語と日本語には似た面があり、これらとフランス語とは著しい対照をなす。このように異なった構造を同一のプログラムで処理するためには、各言語に通じた一つの抽象的なモデルを考えることが有益であろう。すなわち支配項 (GOV) と被支配項 (DEP) は一般に

左 DEP₁ 左 DEP₂ … 左 DEP_n GOV 右 DEP₁ 右 DEP₂ … 右 DEP_n
と考える。これはすでにガイフマンによる定式化がある(文献2)。日本語では GOV の右側の DEP はすべてファイとなる、という特性をもっていると考えるのである。この種の考え方は、言語構造の相互比較、すなわち対照言語学の研究にとって有益であろう。従来言語の対照研究が進まなかったのは、その方法の発見と設定が必ずしも容易でない点にあったといえるのではないか。したがって、そのような比較の方法を作るといふこと自身は、有意義であると考え

えられる。ただし現在のプログラムで、それがどの程度まで可能であるかということは、また別の問題であるが。

このプログラムの特徴の一つは述語をまずみつけ、それを中心として被支配項を順次にみつけていく、というところにある。ただし、それを見つける過程を節約するために、言語を語順のタイプによって三種に分けた。すなわち、主語 S、目的語など O、動詞を V とすると、順列法としては SOV、OSV、SVO、OVS、VSO、VOS の 6 種がありうる。ただしこのうち OSV、OVS、VOS の型を原則とする言語はなさそうである。そこで、実際上は SOV、SVO、VSO の三種あればよいことになる。そこでこの三種の型のなかでそれぞれの述語の見つけ方を考えた。この点、実はプログラムに多少 ununiversal でない点をつくったところがある。

このような述語中心のアルゴリズムでは、限界があるが、さきにもふれたように、言語行動あるいは言語活動のシミュレーションに通ずる道もあると考えている。このためには、話手、聞手、話題、場面などディスコース・アナリシスに通じる多くの点を取り扱えなければならない。一つの文の取り扱っただけでは不十分である。これからの努力の目標としなければならない。

このプログラムは以上のように本来構文の解析を行うプログラムであるが、構文の解析が言語処理の基本問題であるところから、いくつかの応用が考えられる。同音語、同形語を文脈のなかで処理するところから、語彙調査における同語異語の判別、かな入力漢字出力としてのかな漢字変換に応用でき、さらに機械翻訳、質問解答などの言語処理における基礎段階での使用など、さまざまな応用範囲が考えられる。実験としては、同語異語の判別、かな漢字変換における同音語の判定などを行った。

このプログラムは、あまり多くの時間の得られない環境のなかで、急拠企画、設計と製作を行った。そのため、当面の実験に差し支えない範疇で完成しているが、なお手直したい点がある。特にメモリーの大きさを、国語研の HITAC 8250 の第一領域 4 万バイトに合わせて作ってあるため、文の

長さの制限, 同音語の数の制限, 辞書のエントリーの制限, 変形規則の長さの制限など, 各種の制限がある。また形態論的な処理もしていない。将来のことをいえば, 述語中心でなく, string の左端からの分析を考えてみたい (言語行動のシミュレーションとしてはその方が望ましい)。このような改変は, さらに多くの言語における分析をしながら, 行いたいと考えているので, 少々時間を要することとなろう。

このプログラム IRIS0A は HITAC 8250 を使用して実行した。プログラム作成に使用した言語は COBOL であり, 使用エリアは 4 万バイト弱である。

1. プログラムの内容

1.1 入力と出力

初期にパラメータを入力し, 次にテストするセンテンスを一つずつ入力する。一文ずつ解析を行い, 構文解析処理の要点をラインプリンタでメモとして出力する。

初期の入力としては, 語順情報 (WORD ORDER TYPE), 辞書 (LEXICON), 句構造規則 (PHRASE STRUCTURE RULES), 変形規則 (TRANSFORMATIONAL RULES), 変形逆探知テーブル (T-REVERSE), 一般構文規則 (SYNTACTIC RULES. 副詞は動詞にかかる, など結合がやや自由な構文規則), 不定代名詞を指示することば (たとえば semantic feature に +loc とあれば「どこかに」を選び出して, 出力文に補うための代名詞などによる代置語) の表。これらの内容と組み合わせについては 1.2 CARD-READ のところで説明する。

出力としては, テスト文全形 (TEST SENTENCE), 辞書の検索 (LEXICON SEARCH), 同音語があるとき, すべての組み合わせ (COMBI OF INFO), 述語探索の結果 (MAIN PREDICATE VERB), 支配語から, パタンを通して被支配項を捜す (DEPENDENT-SEARCH), 処理されなかった被支配項から支配項をさがす (GOVERNOR SEARCH), 全体のツリーの構造

を一覧する (ANAPHORIC ANALYSIS) などである。

1.2 プログラムの各段階

1.2.1 CARD-READ

各言語のパラメータとしてその解析に必要な事項を含む語彙文法上の記述をカードにさん孔して読みこませる。文法は各種のバラエティをもった記述を許しているのので、読みこみに当ってどの種類のカードがあるとかないとかのチェックはいっさい行わない。カードの種類は

WORD ORDER TYPE VSO, SVO, SOV の三者の選択。もしこれを入れなければ、プログラムは SOV と解釈する。SOV というカードを入れても同じ結果になる。

LEXICON これはアルゴリズム上必須。もし入れないと、辞書を検索してもないというメッセージを出して、LEXICON SEARCH の途中でとまる。品詞は必須、あと必要に応じて文型番号、統語論上の情報、意味論上の情報を書く。

PHRASE STRUCTURE RULES このカードは任意。ラインプリンタ出力中は S-PAT すなわち SENTENCE PATTERN (文型) と印字する。文型番号、支配項 (GOV) の番号、文型中の各項に品詞、統語論上の情報、意味論上の情報などを書く。

TRANSFORMATIONAL RULES このカードは任意。変形規則の番号、変形前の支配項・変形後の支配項の番号、矢印をはさんで変形前諸項と変形後諸項。

T-REVERS 変形逆探知テーブルのことで、このカードは任意。逆探知のばあいの前後にあるはずの項の内容ともどるべき変形番号。

SYNTACTIC RULES ある項目について後または前にどんな項があればその支配項であるかを書く。このカード任意。

INDEF-PRONOUN 解析のあと文型上うまっていない項があり、かつこのカードがあれば、意味論上の情報を通じて出力中に、ナニカ、ドコカ、イツカ、

ETWAS, JEMAND などを補うことができる。このカード任意。

1.2.2 TEST SENTENCE

テストする文を一文ずつ読ませる。文が長ければ2枚にわたってもよいが、処理エリアの都合上、15語をこえると解析をうちきる。

1.2.3 LEXICON SEARCH

テスト文中に含まれる語について辞書に記載されている情報を求める。辞書を引くとき、一つの情報があればそれで満足して次に進むのではなく、辞書に記載されている同意語すべての情報をうつしおえてから次の語の検索にとりかかる。これは同音語処理を完全に行うためである。辞書検索上は形態論的な処理は行わない。

1.2.4 COMBINATION OF INFO

上記の多義語あるいは同音語のすべての組み合わせの文字列を作成する。いま文がABCの三語から成り、Aに2、Bに3個の同音語があるとすれば、組み合わせの総数は6個となる。この場合のつくり方は次のようにする。

A ₁	B ₁	C
A ₂	B ₁	C
A ₁	B ₂	C
A ₂	B ₂	C
A ₁	B ₃	C
A ₂	B ₃	C

このようにしてあらかじめすべての組み合わせをつくって一つずつ磁気テープ上のファイルに記録する。磁気テープへの記録が終了したらファイルをクローズして次のステップ MAIN PREDICATE をさがすルーチンへ移行する。磁気テープへの記録を行うのは、辞書検索のために利用する

(文の長さ) × (同音語の数) × (単語情報の長さ) 単位バイト

のような大きなエリアに収めた情報を上記のようなストリングに分解して磁気テープ上に記録したのち、このエリアをクリアして、一つ一つのストリングの

解析の記録 (S-PAT やそれを变形したものをストアする) をとるために用いるのである。計算機が十分大きければ、エリアを二つ別にとって、磁気テープのファイルをつくらぬ方がよいと思われる。

1.2.5 MAIN PREDICATE

テスト文中の各語につけられた情報 (辞書から書きぬいたもの) をもとにして主な述語を決定するルーチン。ここでは語順によってアルゴリズムを変えてある。VSO 型は文の左端またはそれに近い述語候補語を述語とする。SOV 型では文の右端またはそれに近い述語候補語を述語とする。SVO 型では左から見ていって文の左端に近い述語候補語をとるが、もしそれまでに句をひきいる語 (関係代名詞、関係副詞その他) があれば、その句の述語をぬかして次の述語候補語をとる。このようにして得られた述語はその文のなかでの語のナンバーをストアして、解析などの出発点とする。

1.2.6 DEPENDENT SEARCH

まずサブルーチンの文型サーチを利用して文型を求める。文型を求めたら、そのなかにしるされている被支配項 (DEP) を文中のなかにさがす (パターン・マッチング)。マッチングができたものについて記録をとる。この際、支配項 (GOV) から DEP への方は変更しない。マッチングに当たっては、品詞、統語情報、意味情報を比較する。比較に際してはパターン側に記入があればそれを重視してテスト文の語にそのフィーチャーがあるかどうかを見る。記入がなく空白になっているときは、基準に合っているものとみなす。

文型の上でははじめから文型を変えることを要求しているもの (たとえばラレルを伴うばあいなど) はそれを実行してからパターン・マッチングを行う。また文型をさがしても得られないもの、文型をさがすように書いてないものについては、このルーチンをパスして次のルーチンに移行する。

一つの文型とテスト文のパターン・マッチングが終了すると、そこで得られた被支配項の一つを次の GOV と見なして文型を求め、その DEP を探す。すなわちこのルーチンのはじめにもどって同じ操作をくりかえす。もし新しい DEP

が何もえられないレベルに到達したら、このルーチンをぬけて次のルーチンに移行する。

1.2.7 GOVERNMENT SEARCH

前項とちがって DEP をとりあげて GOV を探すルーチンであって、いまだに帰属がきまっていない語を一つ一つ取りあげてその GOV を捜してゆく。すべての語についてその GOV がわかったら、このルーチンをぬけて次の ANAPHORIC ANALYSIS に移る。この GOVERNMENT SEARCH は次の二つの部分に分かれる。

1.2.7.1 変形逆探知

解決のつかない語を中心にその近辺の構造を調べてなにか文型に変形がかかっているのではないかと捜す。その可能性があれば、必要があればその文型を求め、それに逆探知した変形を施して新たな文型とみなし、GOV に異動があるかどうかを調べ、必要があれば新しい GOV を見定めて、1.2.6 の DEPENDENT SEARCH に移る。この際従来行った解析と交差抵触をおこす可能性のあるものについては、その部分だけ従来の解析記録を抹消してから前のルーチンにもどる。この 1.2.7.1 のルーチンで逆探知ができないもの、(はじめからないものも含めて) 文型のないものは、次の 1.2.7.2 に移る。

1.2.7.2 品詞だけの情報による処理

統語論的な情報(たとえば性・数)、意味論的な情報(+animate など)の情報は一切ふれず、もっぱら品詞だけで切れ続きを定める。DEP から GOV への続き方をきめるテーブルを入れておけばよい。もしこのルーチンを通してもまだ帰属のきまらない辞項があれば解析(PARSING)を GIVE UP する。

1.2.8 ANAPHORIC ANALYSIS

はじめにリザーブした MAIN PREDICATE から出発して入力文の tree-structure を書く。文型(S-PAT)があればそれも印字し、そこに求められた DEP をならべてゆく。文型にあって該当する項がなく、かつ INDEF-PRON が読みこんである場合は意味情報を利用しつつ代置語であなうめする。これに

は文中の語のナンバリング(表層構造)を示すことはできないから、それが深層構造ではあったが表層で消えたものと解釈することを許すものである。GOVの記載だけがあってDEPがないのは、その語にかかってくる要素がないことを示す。このようにしてすべての tree structure を書きおえると次の string を読むために 1.2.4 の出力結果を読みこんで 1.2.5 の処理に移ることになる。この磁気テープ上の記録をすべて調べつくした場合は 1.2.2 の次の TEST SENTENCE の読みこみに移る。すべての文を処理すると JOB は完了する。

以下は上記のルーチンのなかで使用するサブルーチンである。サブルーチンのなかで他のサブルーチンを利用しあうこともある。

1.2.9 S-PATTERN SEARCH

辞書に記載されている単語の文型番号を使って実際にその文型を求める短いルーチン。このサブルーチンは COBOL の PERFORM 動詞を用いてメインまたはサブルーチンのなかで用いられる。

1.2.10 TRANSFORMATION

辞書に記載されている変型番号または逆探知された変型番号を使用してその変形の式を求め、変形を受ける文型を変形するルーチン。変形の式の各項には矢印の左側にも右側にもすべて番号が与えられている。矢印の左側にある番号で右側になければ削除、矢印の左側になくて右側にあるものは挿入、矢印の左側と右側で番号の順序に相違があるときは順序の変更という操作を施すことになる。このルーチンはまず矢印の左側の項と変形を受ける文型の項を比較して対応するものがあれば変形式の番号を文型の項に移し、次に矢印の左側の並べ方に従って変形を受ける文型の項を変形を受けた文型(の項をならべるところ)に順次に移す。この際言語による WORD ORDER TYPE により、変形を受ける文型中の番号のない項は変形を受けた文型のどこかに入れる(たとえば日本語的な語順では GOV の左側であればどこへ入れてもよいと考えられるので、とりあえず一番はじめに移動する)。このルーチンはかなり長いが COBOL の PERFORM 動詞を用いてメインまたはサブルーチンのなかで用いられる。

2. 解析実験とその情報科学および言語学的な問題点

2.1 行った実験

時間的な余裕がそれほど多くはなかったので、あまり多くの実験を行うことはできなかった。言語も日本語、ドイツ語、フランス語に限られている。このなかでは現代の日本語を主として（なかに平安朝の「竹取物語」の例を少し入れた）取りあげた。材料は漱石鷗外などの文学作品、現在行っている教科書用語調査などから取った。文法理論はかかり受けを主とする伝統文法、変形規則を別に添えた、意味特徴を書きこんだ結合価文法を用いている。フランス語、ドイツ語については、変形規則を別に添えた、意味特徴を書きこんだ結合価文法を用いた。日本語については、応用例として語彙調査における同語異語の判別およびかな入力漢字出力のコンピュータ内での同音語判別の部分の実験も行った。解析のいくつかの例としてラインプリンタ出力の例をうしろに示す。

2.2 言語的な問題点

最大の問題点は句構造規則と変形、あるいは意味特徴を書きこんだ結合価文法と変形規則の組み合わせで、すべての文の解析が可能であるかどうかという点にある。理論的な解明が最も重要であることはいうまでもないが、この種の実験を通してそれがどこまで可能であるかをつきとめることも必要であろう。実験を行うことによって問題を具体的にとらえることができると考えている。しかしこのためには多くの言語の多くのデータについて実験を行う必要がある。

このことに関してはベヴァーなどの perceptual strategy というような仮説が提出されている(文献3)。その線に沿った研究も進んでいる模様である。このためには、特に心理言語学ないしは心理学的な研究を進めることが必要であろう。それをアルゴリズム化し、プログラム化するためには、全体としてある程度見通しがついていることが必要である。それは今すぐ始められることであるのかもしれないが、しかし研究の段階がその段階であるなら、いまは IRIS0A のアルゴリズムによってそのアプローチでの限界に見通しをつけることもなお

有益であろう。

そのことと、述語中心ということとも関係がないとはいえない。述語を中心とすると、多くの点で人間の理解行為はかなりかけ離れた操作をしなければならなくなる。このプログラムで述語中心という原則を立てたのは、いろいろな点でその方が解析のための文法を書くのに便利ではないかと考えられたからであった。できれば、文頭からの解析にある程度の見通しをつけることが必要であり、それができたら、言語一般にわたる文頭からの解析アルゴリズムによるプログラムを作成する必要があると考える。その方が人間の理解の行為のモデルとしてはよいものになるだろう。その際にはさきの perceptual strategy あるいはそれに代わるものを、なかに組みこむことは十分検討されてよい。

言語行動のシミュレーションとして考えてみると、このような処理のアルゴリズムは、いわば生体内の情報処理に当たる部分であって、それも、句構造文法と変形規則およびその運用（プログラム）という部分のシミュレーションにすぎない。これはいわば言語内的な処理の部分であり、現実の言語行動のなかの一部である。これを広げるためには言語内の事実から論理・心理的なものへ、さらに具体的、全体的な人間行動のなかでの位置とが考えられなくてはならない。これらの諸過程のなかにプラグマティズム、話手、聞手、場面、社会、それらの関係などが定位されなければならない。このプログラムでは言語処理上の必要から、理解の処理を中心としているが、表現の側からのアプローチが必要であることはいうまでもない。言語のモデルとしてどのような条件があるか、生成の過程としてどんな問題があるかを明らかにしなければならない（文献4）。言語内と言語外との間にどのようなリンクージュをとるか、この点だけでもさまざまなモデルが考えられよう。

以上のことと関係するが、このプログラムが同一のプログラムで多くの言語を処理するという、そのために句構造規則や変形規則などを利用し、各言語の特殊性はこれらのなかに記述されるということから、逆にこのことを利用して、諸言語間の構造の比較のための用具として使用することも考えられる。

これまでいわゆる対照言語学が提唱されてきたが、その方法がなかなか具体化しなかった。最近になって生成文法の手法を取り入れて、比較的進展しつつある。比較対照のためには、なにかの点で共通の基礎に立っていることが必要である。その意味で句構造規則や変形規則の運用が同一のアルゴリズムの上に行っているということは有効なのではないかと考えている。ただしその比較は、もちろんさきにもべた通りの、いわば言語内的な部分に限られてしまうので、本当の意味での対照言語学は、それを含みながらさらに広い部分もおおうべきものであると考えるとすると、これでは十分とはいえない。この意味で対照言語学にとっても、言語の全体的なモデルを取りあげることが必要であろう。

プログラム IRIS0A 作成の目的の一つはさまざまな文法理論がどのような射程をもつかということであり、その比較であった。実験では「コンナ ユメ ヲ ミタ」について、かかり受けを重視する伝統的な文法にしたがって、コンナはユメにかかりユメはヨにかかり、ヨはミタにかかるという処理ができた。これは一つの入子型の処理とみてもよい。これを結合価文法で処理すると、ユメ ヲがミタにかかることはそれでよいのであるが、「ダレカ ガ」という部分が、入力した表層構造には欠けている。そこでプログラムは、「ダレカ ガ」を補っている。純粹なかかりうけでは、そのような配慮は必要ではないから、できなくてもよいのであるが、この種の自動処理では、深層構造にはあるはずであるが、表層にはないところの「ダレカ ガ」が指摘されていることは、きわめて有意義である。その意味で、言語処理には、伝統的な国文法よりも、結合価文法の方がより有効であるということが出来る。(もっとも、これを延長して直ちに、言語処理以外に拡張して「あらゆる場合に」というつもりはない。)さらに、変形規則を適用して(しかもくりかえし適用して)結合価のさまざまなヴァリエーションに即応しうることは、現実の複雑な表現の分析に有利であることはいうまでもない。

IRIS0A で「変形」といっているものは、現在のいわゆる変形文法とは、その考え方に少し相違がある。そして、IRIS0Aの方が広く、かつその記述が記

述者の任意にまかされているのであるから、いわゆる変形文法の範囲にせまく限定して使用すれば、実質上差はないことになる。IRIS0A では、とにかく記述者が制限をつけなければ（ある変形規則は文型何番から何番までの間に限って働かせることができる、というような制限が使えるようになっているが）、あらゆる文型にあらゆる形で適用される。そこで、現在日本語について日本の変形生成文法の人々が必ずしも用いていないような使いかたをすることができるし、またある場合には、それが必要である。ある場合というのは、特に「埋めこみ文」がある場合であって、いわゆる生成の場合には、学者がいきなり「埋めこみ文」の存在とその形を提示してみせるが、解析の機械処理のばあいには解析を始める前からうめこみ文の存在とその形を予言することは本来不可能である。そこで変形の操作を表層の形から深層の形を推測するために及ぼすようなものにまで広げることも許されるはずである。筆者は在仏時代、ニコラ・リュヴェ氏と話す機会を得たとき、このことについてふれたが、リュヴェ氏は、それはそういう必要があるのだから、その立場からは認められるべきだといわれた。そこで、「タカセブネ ハ キヨウト ノ タカセガワ ヲ ジョウゲスル コブネ デ アル」というような文の処理にあたって、「フネ（コブネ）ガ カワ ヲ ジョウゲスル」というパターンを設定しておき、「ジョウゲスル（コ）ブネ」のような連体形のパターンを入れておかず、「名詞 ガ 動詞」という文型を「動詞 名詞」に変えるような変形規則を入れておけば、コブネにかかるものとしてキヨウト ノ タカセガワ ヲ ジョウゲスルが表層であるように指摘するようになる（文献5）。

2.2 情報科学としての問題点

情報科学の問題点としても最大の問題点は、このプログラムが言語内容的な操作にとどまっている点であり、言語モデルのごく一部をねらったという限定であろう。言語学的な問題点のところでも述べたことと全く同一のことが、ここでもいえる。そういうインテグラルな態度が今度のはのぞまれもしようし、ディコースをどう扱うもかそれができなければならない（文献6, 7, 8）。京大の

PLATON ではすでに back tracking の機能を利用して、ある程度文をこえた処理も可能なようである。

IRIS0Aの特徴は変形逆探知を統一したやり方でできることであって、前述の「変形」を拡張したために可能になったのである。しかしこのような措置がどこまで可能か、理論的にまた実験を通してもっとよく確かめることが必要である。

ディスコース・アナリシスへのアプローチとしては、表層にない要素の推定（ダレカガ、ドコカデなど）は日本語の場合その第一歩であろうと思う。そのだれか、どこかが具体的にだれであり、どこであるかは、場合によっては同一文中に求めることもできるが、多くのケースでは、もっと長い文脈の検討が必要となるであろう。その辺から始まって、文と文とのつながり、段階としてのまとまりが求められることになろう。筆者としては、このプログラムですでに始めたコンピュータ言語学的なディスコース・アナリシスへのアプローチを、論理的な分析への志向において発展させるのが適当ではないかと考えている。その場合、取り扱う対象としてはもちろん、論理的な文章すなわち科学技術の文献を考えている。

プログラム作成技術としては、プログラムが長くなっていること、エリアのとり方に多少むだがあることが気になる。途中で磁気テープを使っているのは解析を遅くしている原因になっている。磁気テープの使用については、一文の解析中に出力し次に入力するというのを一回は必ずすることになっており、単語について多義語、同形語の取り扱いを前述のように行うことから、一文の中の単語の最大数の数のそれぞれに同形語数を示すインデクスを付して使うためかなり大きなエリアを使っている。磁気テープを使って組み合わせをあらかじめ作っておき、磁気テープ記入後はこの大きなエリアを直ちにクリアして解析のための文型を求めたり、その変形を行ったり、DEPENDENT, GOVERNOR を書きこんだりするために使用しているのだから、磁気テープをなくすとそのエリア（このプログラムで使用するエリアのうち最大のもの）を二倍にしなければならぬ。

プログラム作成者にとって、作成上最大の問題点はプログラムのエリアであった。国立国語研究所のコンピュータは98Kバイトのメイン・メモリーで、ユーザース・エリアは 70 KB、これを通常 40 KB, 30 KB の二つの領域に分けて使用している。用語用字調査のための使用時間が多く、第二領域はそれで使用することが多いので、IRIS0A は第一領域 40 KB に収めることとし、磁気テープ入出力の オールターナティブ・エリア を節約、辞書・文法規則などを縮小、チェックルーチン等を割愛その他いくつかの機能をあえて削除して、実験プログラムとして最小限の機能を確保し、ようやく 40 KB すれすれのところで出来上った。

これからのプログラム上の課題は決して少なくない。できれば磁気テープ・ファイルをなくすこと(磁気ディスク・ファイルを使用すればよいのであるが、それでは LABEL RECORD OMITTED を使用することができず、ラベルを STANDARD にすると自動処理中しばしば、すなわち一文処理中必ず一回は日付けのチェックで止まるおそれがある。磁気テープは LABEL RECORD OMITTED が可能)、辞書・文法規則は磁気ディスクに収めること、プログラムのオーバーレイを行ってエリアの縮少をはかり、必要な機能を増加すること、などであるが、それにも増して重要なのは、いろいろな、性質の異なった言語の、さまざまな構造の文を解析してみて、このプログラムのアルゴリズムを徹底的にチェックし、全体的な補正を行うことである。オーバーレイの実行などの修正はそれからあとでよいと思う。

3. 名称, 引用文献, 流れ図, 入力出力の見本など

プログラム名称 IRIS0A の IRIS はもと文部省科学研究費「作品の用語検索組織の研究」(研究代表者石綿敏雄)を得て作成した KWIC 作成などのプログラム IRIS と共通の名であり、その方の初めの I は作成者のイニシャル(国語研内の約束による)あとの R は caRd, I は Input, S は System であるが、全体の名はギリシャ神話の「虹」の女神になっている。IRIS 01~IRIS 09 は

同語異語判別をカードによるターン・アラウンド・システムを含めて行うカード入力による KWIC 作成, 用語用例情報検索システムである。IRIS0A が構文解析のプログラムで, カード入力は両者共通であるから, データ上の相互交換が可能である。

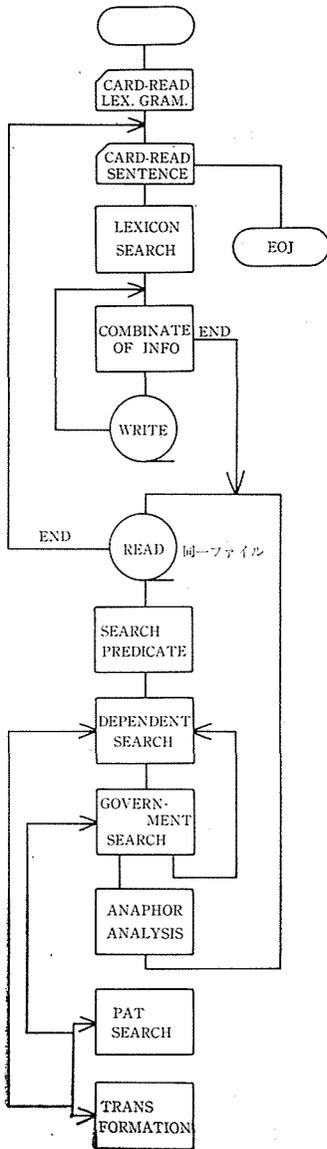
【参考文献】

1. 石綿敏雄「日本語の生成語彙論的記述と言語処理への応用」国語研報告54『電子計算機による国語研究』Ⅶ 1975 所収
2. B. VAUQUOIS “La traduction automatique en Grenoble” 1974
3. 石綿敏雄「計算言語学最近の動向」『月刊言語』 1974・9
4. 南不二男「現代日本語の構造」 1974
5. 石綿敏雄「コンピュータによる言語処理」『日本のことばと文化・社会』 第五巻 所収
6. 中井浩・高野守正「論理と情報の世界」1970
7. 長尾真・辻井潤一・田中一敏「意味および文脈情報を用いた日本語文の解析——文脈を考慮した処理」『情報処理』 Jan. 1976
8. 林四郎「文の姿勢の研究」 1973

【流れ図および入力・出力見本】

- ① INPUT SENTENCE 見本。
- ② 教科書調査「地学」のテキストから。DEPENDENT-SEARCH の GOV S-PATTERN 08 は述語動詞の文中の単語番号。SEMANTIC-LEVEL 08 02 とは 02 の「ニ」が 08「アル」にかかるの意。ANAPHORIC ANALYSIS の GOV 08 に DEP 06, 07, 01, 02, 03 がそれぞれかかり, 06 には 04, 05 がかかっていることを示す。GOV のみあって DEP のないのはそれに続く要素のないことを示す。
- ③④教科書調査「物理」のテキストから。接続助詞の処理は二つの動詞をつなぐコネクタのように考えて行える。大きな家庭用電気製品にさらに電気のとりにちがついているのに似た処理法。
- ⑤ 語彙調査における 同語異語判別への応用例。「が」は格助詞 (IC) と接続助詞 (CJ) とがあるので, まず COMBINATION OF INFO でそれぞれの解をとる二つの strings のありうることを示し, それぞれをとりあげる。そして格助詞の解では統語論的な検討がうまくいくが, 接続助詞という解をとると, 「が」の直前に名詞があるので, 解析を GIVE UP する。したがって格助詞という解だけが検討をパスしたことになる。

- ⑥ かな漢字変換への応用例。「ジカン」が「時間」(time)と「次官」(viceminister)の同音語。これを見わけようとする。このばあいには「出席する」という動詞の文型で「が」の前が意味特徴として +hum. をもつものであることを利用し、「カイギ ニ ジカン ガ シュッセキスル」では、「時間」という解釈はできず (GIVE UP PARSING 03), 「次官」なら可能という判定を行っている。
- ⑦ かな漢字変換への応用例の一。「ジカン ヲ タンシュクスル」では「ジカン」は「次官」で解釈することはできず、「時間」のばあい可能であるという判定をしている。
- ⑧ かかり受けの情報を与えただけでプログラムを動かした例。文型を与えないので DEPENDENT SEARCH は素通り、GOVERNOR SEARCH のところで SYNT (ACTIC)—REL (ATION) だけで処理。
- ⑨ 上と同じ例を結合儒文法で処理した例。ANAPHORIC-ANALYSIS で、グレカガ を補っている。
- ⑩ 文型には「アメ ガ フル」とあるのに、TEST SENTENCE は「アメ ニ フラレル」である。プログラムは「フラレル」に注目してまず変形を行い、次にモデルと入力文のパタン・マッチを行う。
- ⑪⑫ ドイツ文 "FRAGEN SIE SPAETER EINMAL IHRE SCHWESTER" を解析する。まず SCHWESTER につき Nominativ (1) と Akkusativ (4) の解があることとし、それぞれにつき検討、まず Akkusativ として解析を始めるが、FRAGEN の文型は「Nominativ の名詞+動詞+ Akkusativ で人の意味のある名詞」であり、SCHWESTER のグループは解析可能であるが SIE (02) がうまく解析できずにのこる。そこでこの近所の文脈をみると FRAGEN という動詞が前にあり、自分自身は Nominativ である。一般に「Nominativ 名詞+動詞」は疑問や SIE への命令のときは順序が変わる規則があることを探り出し、文型を変形してあらためてパタンマッチングを行う。この解析は成功するが、SCHWESTER を Nominativ とすると解析ができない。
- ⑬ ドイツ文で定動詞と組み合わせる不定法は文末にくる。この場合でも文型は「定動詞+不定法」、「形容詞など+不定法動詞 (VI)」でうまく処理できる。
- ⑭ フランス語で関係代名詞をうちにもつ構造の文の解析。
- ⑮ 入力フランス文に誤りのある例の処理。フランス語の名詞句のなかでは形容詞、冠詞が名詞と性・数という文法的なカテゴリーで一致するということが書かれているので、その場所(06)を指摘して解析をやめる。文法的な誤りを正しく直すことはできない。
- ⑯ 同上のフランス文の誤りをただし、誤りのない文にして入力した例。機械は最後まで解析を行い、完了する。



FLOW CHART

of an automatic syntactic analysis system which treats transformational rules, made by T.ISHIWATA

⑤

TEST SENTENCE
 LEXICON SEARCH

01	530012	N	H	530012
02	#	1C	SH	#
03	#	CJ	CJ	#
04	#	N	CDDY	#
05	#	2C	LD	#
06	#	VW	LD	#

COMBINATION OF INFO

01	530012	N	H	530012
02	#	1C	SH	#
03	#	N	CDDY	#
04	#	2C	LD	#
05	#	VW	LD	#

FOR A STRING SUCH AS

01	530012	N	H	530012
02	#	1C	SH	#
03	#	N	CDDY	#
04	#	2C	LD	#
05	#	VW	LD	#

MAIN PREDICATE VERB
 DEPENDENT-SEARCH

03	1105	N	H B 1C	SUB N	OYB 2C	LOB VN	B
----	------	---	--------	-------	--------	--------	---

SEMANTIC-LEVEL
 SEMANTIC-LEVEL
 GOVERNOR-SEARCH
 SEMANTIC-LEVEL
 ANAPHORIC-ANALYSIS

05	#	VW	LD	#
06	#	VW	LD	#

S-PAT

01	530012	N	H	530012
02	#	1C	SH	#
03	#	N	CDDY	#
04	#	2C	LD	#
05	#	VW	LD	#

FOR A STRING SUCH AS

01	530012	N	H	530012
02	#	1C	SH	#
03	#	N	CDDY	#
04	#	2C	LD	#
05	#	VW	LD	#

MAIN PREDICATE VERB
 DEPENDENT-SEARCH

03	1105	N	H B 1C	SUB N	OYB 2C	LOB VN	B
----	------	---	--------	-------	--------	--------	---

SEMANTIC-LEVEL
 SEMANTIC-LEVEL
 GOVERNOR-SEARCH
 GIVE UP PARSING

⑦

TEST SENTENCE
 LEXICON SEARCH

01	530012	NH	TH	E601 TIME
02	#	3C	HU	E105 VICE-MINISTER
03	5300226	VW		

COMBINATION OF INFO

01	530012	NH	TH	E601 TIME
02	#	3C	HU	E105 VICE-MINISTER
03	5300226	VW		

FOR A STRING SUCH AS

01	530012	NH	TH	E601 TIME
02	#	3C	HU	E105 VICE-MINISTER
03	5300226	VW		

MAIN PREDICATE VERB
 DEPENDENT-SEARCH

03	0203	NH	T'B 3C	B VN	B
----	------	----	--------	------	---

SEMANTIC-LEVEL
 SEMANTIC-LEVEL
 GOVERNOR-SEARCH
 ANAPHORIC-ANALYSIS

03	5300226	VW			02
04	5300226	VW			02

S-PAT

01	530012	NH	TH	E601 TIME	
02	#	3C	HU	E601 TIME	
03	5300226	VW			02

FOR A STRING SUCH AS

01	530012	NH	HU	E105 VICE-MINISTER	
02	#	3C	HU	E105 VICE-MINISTER	
03	5300226	VW			02

MAIN PREDICATE VERB
 DEPENDENT-SEARCH

03	0203	NH	T'B 3C	B VN	B
----	------	----	--------	------	---

SEMANTIC-LEVEL
 GOVERNOR-SEARCH
 GIVE UP PARSING

TEST SENTENCE # 307 1/ 1 2/ , #

LEXICON SEARCH 01 307 RN NJ
02 2/ NJ
03 2/ JV
04 2/ VVA
05 ,

COMBINATION OF INFO 307 RN
2/ NJ
2/ JV
2/ VVA
,

FOR A STRING SUCH AS 01 307 RN
02 2/ NJ
03 2/ JV
04 2/ VVA
05 ,

MAIN PREDICATE VERB 04
DEPENDENT-SEARCH
GOVERNOR-SEARCH
SYNT-REL 04 03
SYNT-REL 03 02
SYNT-REL 02 01
SYNT-REL 04 03 05

ANAPHORIC-ANALYSIS
GOV 04 2/ VVA
DEP 03 2/ JV
DEP 05 ,
GOV 03 2/ JV
DEP 02 2/ JV
GOV 05 1/ NJ
DEP 02 1/ NJ
GOV 01 01 307 RN
307 RN

TEST SENTENCE # 0105 1/ 1 2/ , #

LEXICON SEARCH 01 0105 RN AC
02 = 2C
03 1/ RN TH E461 TIME
04 1/ RN HU E105 VICE-MINISTER
05 0207020 VVA 01

COMBINATION OF INFO 0105 RN AC
= 2C
1/ RN TH E461 TIME
1/ RN HU E105 VICE-MINISTER
0207020 VVA 01

FOR A STRING SUCH AS 01 0105 RN AC
02 = 2C
03 1/ RN TH E461 TIME
04 1/ RN HU E105 VICE-MINISTER
05 0207020 VVA 01

MAIN PREDICATE VERB 05
DEPENDENT-SEARCH
GOV S-PATTERN 05 0105 RN HUB IC B SM ACS 2C B VV B
SEMANTIC-LEVEL 05 02
SEMANTIC-LEVEL 05 01
SEMANTIC-LEVEL 05 04
GOVERNOR-SEARCH
GIVE UP PARSING 03
FOR A STRING SUCH AS 01 0105 RN AC
02 = 2C
03 1/ RN HU E105 VICE-MINISTER
04 1/ RN HU E105 VICE-MINISTER
05 0207020 VVA 01

MAIN PREDICATE VERB 05
DEPENDENT-SEARCH
GOV S-PATTERN 05 0105 RN HUB IC B SM ACS 2C B VV B
SEMANTIC-LEVEL 05 02
SEMANTIC-LEVEL 05 01
SEMANTIC-LEVEL 05 04
SEMANTIC-LEVEL 05 03
GOVERNOR-SEARCH
ANAPHORIC-ANALYSIS
GOV 05 0207020 VVA 01
S-PAT 05 0105 RN HUB IC B SM ACS 2C B VV B
DEP 03 03 04 SM 01 SM 02 SM 03 SM
DEP 04 1/ RN HU E105 VICE-MINISTER
DEP 01 0105 RN AC
DEP 02 1/ RN HU E105 VICE-MINISTER
GOV 03 03 04 SM 01 SM 02 SM 03 SM
GOV 04 1/ RN HU E105 VICE-MINISTER
GOV 01 0105 RN AC

GOV 02 1 2C

TEST SENTENCE # 307 2X X 29 . X

LEXICON SEARCH
 01 307 R N 307
 02 2X N N 307
 03 2X N N 307
 04 29 VVN VVN 05 2X 9
 05 29 VVN VVN 05 2X 9

COMBINATION OF INFO
 307 R N 307
 2X N N 307
 2X N N 307
 29 VVN VVN 05 2X 9
 307 N N 307
 2X N N 307
 29 VVN VVN 0205 2X 9

FOR A STRING SUCH AS
 01 307 R N 307
 02 2X N N 307
 03 2X N N 307
 04 29 VVN VVN 05 2X 9
 05

MAIN PREDICATE VERB
 DEPENDENT-SEARCH
 GOV S-PATTERN 04 0995 N H B IC SUB N A1B 3C DBB VN 8
 SEMANTIC-LEVEL 04 03
 GOVERNOR-SEARCH 04 02
 S-VT-NEL 04 03
 SYL-NEL 04 05

ANAPHORIC-ANALYSIS
 GOV 04 29 VVN 05 2X 9
 S-PAT 04 0995 N H B IC SUB N A1B 3C DBB VN 8
 N 2X 9
 02 2X IC SU N 2X
 03 2X IC DB 2X
 GOV 02 2X N A1 2X
 DEP 03 2X N 3C DB 2X
 GOV 04 2X N 2X
 DEP 01 2X N 2X
 GOV 03 2X N 3C DB 2X
 GOV 04 2X N 2X

FOR A STRING SUCH AS
 01 307 R N 307
 02 2X N N 307
 03 2X N N 307
 04 29 VVN VVN 0105 2X 9
 05

MAIN PREDICATE VERB
 DEPENDENT-SEARCH
 GOV S-PATTERN 04 0505 N H B IC SUB N A1B 3C DBB VN 8
 TRANSFORMED BY 012327 N B11C SUBVN B3-----N B17C SUBVN B3

TEST SENTENCE # 77 = 0500 . #

LEXICON SEARCH
 01 77 N COPH 77
 02 = 7C LU 0201 77 06
 03 700 VVN VVN
 04

COMBINATION OF INFO
 77 N COPH 77
 = 7C LU 0201 77 06
 700 VVN VVN 0201 77 06

FOR A STRING SUCH AS
 01 77 N COPH 77
 02 = 7C LU 0201 77 06
 03 700 VVN VVN
 04

MAIN PREDICATE VERB
 DEPENDENT-SEARCH
 GOV S-PATTERN 04 0117 N CVB IC SUB VN 8
 TRANSFORMED BY 012327 N PHB11C S-P2VN B3-----N B12C B4VN B3
 TRANSFORMED FROM 0105 N CVB IC SUB VN B
 INDEXED FROM 0105 N CVB IC SUB VN B
 TRANSFORMED TO 0105 N CVB12C S4VN B3
 SEMANTIC-LEVEL 01 12
 GOVERNOR-SEARCH 01 11

ANAPHORIC-ANALYSIS
 S-VT-NEL 04 04
 S-PAT 03 700 VVN 0201 77 06
 04 03 N CVB12C S4VN B3
 DEP 01 77 N COPH 77
 DEP 02 = 7C LU 77
 DEP 03 77 N COPH 77
 GOV 01 77 N COPH 77
 GOV 02 = 7C LU 77
 GOV 04 =

(11)

TEST SENTENCE # FRAGEN: SIE SPAETER EINMAL IHRE SCHWESTER #

LEXICON SEARCH

01	FRAGEN	VVD S 2	02
02	SIE	NM S12	
03	SPAETER	AD	
04	EINMAL	AD	
05	IHRE	AJFS4	
06	SCHWESTER	NMFS4 H	10
06	SCHWESTER	NMFS1 H	

COMBINATION OF INFO

	FRAGEN	VVD S 2	02
	SIE	NM S12	
	SPAETER	AD	
	EINMAL	AD	
	IHRE	AJFS4	
	SCHWESTER	NMFS4 H	10

FOR A STRING SUCH AS

01	FRAGEN	VVD S 2	02
02	SIE	NM S12	
03	SPAETER	AD	
04	EINMAL	AD	
05	IHRE	AJFS4	
06	SCHWESTER	NMFS1 H	10

MAIN PREDICATE VERB

DEPENDENT-SEARCH

01	0202	NM 1	B VD	B NM	4 H B
01 06	1003	ARFS4	P AJFS4	P NMFS4	B NR P

SEMANTIC-LEVEL

GOVERNOR-SEARCH

01	0202	NM 1	B VD	B NM	4 H B
----	------	------	------	------	-------

GOV S-PATTERN

TRANSFORMATION

020205	NM 1	B VD	02XX		-----VD	B2NM 1 . B1XX
0202	NM 1	B VD	B NM	4 H B		
0202	NM 1	B VD	B2NM	1 B1NM	4 H B	

INDEXED FROM

TRANSFORMED TO

SEMANTIC-LEVEL

01 02	01	VD	B2NM	1 B1NM	4 H B
-------	----	----	------	--------	-------

SEMANTIC-LEVEL

01 06	1003	ARFS4	P AJFS4	P NMFS4	B NR P
-------	------	-------	---------	---------	--------

GOV S-PATTERN

SEMANTIC-LEVEL

06 05					
-------	--	--	--	--	--

GOVERNOR-SEARCH

SYNT-REL

01	03				
01	03	04			

SYNT-REL

ANAPHORIC-ANALYSIS

01	FRAGEN	VVD S 2	02
S-PAT	01	VD	B2NM 1 B1NM 4 H B

DEP

02	SIE	NM S12	06 54
06	SCHWESTER	NMFS4 H	10
03	SPAETER	AD	
04	EINMAL	AD	
02	SIE	NM S12	

GOV

(12)

TEST SENTENCE # WIR WOLLEN HEUTE NICHT TRAUERIG SEIN #

LEXICON SEARCH

01	WIR	NM P11	
02	WOLLEN	VVD P 1	06
03	HEUTE	AD	
04	NICHT	DA	
05	TRAUERIG	AJ	
06	SEIN	VI	05

COMBINATION OF INFO

	WIR	NM P11	
	WOLLEN	VVD P 1	06
	HEUTE	AD	
	NICHT	DA	
	TRAUERIG	AJ	
	SEIN	VI	05

FOR A STRING SUCH AS

01	WIR	NM P11	
02	WOLLEN	VVD P 1	06
03	HEUTE	AD	
04	NICHT	DA	
05	TRAUERIG	AJ	
06	SEIN	VI	05

MAIN PREDICATE VERB

DEPENDENT-SEARCH

02	0602	NM 1	B VD 1	B VI	P
02 01					
02 06					
06 05	0502	AJ	B VI	B	

SEMANTIC-LEVEL

GOVERNOR-SEARCH

02	03				
05	04				

SYNT-REL

ANAPHORIC-ANALYSIS

02	WOLLEN	VVD P 1	
S-PAT	0602	NM 1 B VD 1 B VI	06 54 P

DEP

01	WIR	NM P11	
06	SEIN	VI	05
03	HEUTE	AD	
01	WIR	NM P11	
06	SEIN	VI	05
05	TRAUERIG	AJ	
03	HEUTE	AD	
05	TRAUERIG	AJ	
04	NICHT	DA	
04	NICHT	DA	

GOV

```

GOV 06 SCH-ESTER HMF54 H 10
S-PAT 1903 APF54 P AJF54 P NMF54 B NR P
DEP 05 IHRE SPACER AJF54
GOV 03 05 IHRE SPACER AD
GOV 04 EIN'NAL AD
GOV 05 IHRE AJF54
FOR A STRING SUCH AS
01 FRAGEN VVO 5 2 02
02 SIE IM SIZ
03 SPACER AD
04 EIN'NAL AD
05 IHRE AJF54
06 SCH-ESTER HMF51 H

MAIN PREDICATE VERB
DEPENDENT-SEARCH
GOV S-PATTERN 01 0202 NM 1 B VD B NM 4 H B
GOVERNOR-SEARCH
GOV S-PATTERN 01 0202 NM 1 B VD B NM 4 H B
TRANSFORMATION
TRANSFORMED BY 020205 NM 1 B VD B2X -----VD B2M 1 B1XK
TRANSFORMED FROM 0202 NM 1 B VD B NM 4 H B
INDEXED FROM 0202 NM 1 B1VD B2M 4 H B
TRANSFORMED TO 01 02 01 VD B2M 1 B1M 4 H B
SEMANTIC-LEVEL
GOVERNOR-SEARCH 01 01 03
SYNT-REL 01 03
SYNT-REL 01 03
GIVE UP SEARCHING 05 03 04

```

```

TEST SENTENCE
LEXICON SEARCH # LF GARCUN QUI VA A L' ECOLE EST BEAU #
01 LF ARMS
02 GARCUN NMMS HU 11
03 QUI S'R 14
04 VA VVO 06
05 A RE 09
06 L' ARFS
07 ECOLE NMFS LD 10
08 EST VVO 05
09 BEAU AJMS

COMBINATION OF INFO
LF ARMS
GARCUN NMMS HU 11
QUI S'R 14
VA VVO 06
A RE 09
L' ARFS
ECOLE NMFS LD 10
EST VVO 05
BEAU AJMS

FOR A STRING SUCH AS
01 LF ARMS
02 GARCUN NMMS HU 11
03 QUI S'R 14
04 VA VVO 06
05 A RE 09
06 L' ARFS
07 ECOLE NMFS LD 10
08 EST VVO 05
09 BEAU AJMS

MAIN PREDICATE VERB
DEPENDENT-SEARCH
GOV S-PATTERN 01 0502 NMMS B VD B AJMS B
SEMANTIC-LEVEL 01 02
GOV S-PATTERN 02 1101 ARMS P AJMS P NMMS B NR P
SEMANTIC-LEVEL 02 01
GOV S-PATTERN 03 1401 NR B VD B
SEMANTIC-LEVEL 03 04
GOV S-PATTERN 04 0601 VD B RF B
SEMANTIC-LEVEL 04 05
GOV S-PATTERN 05 0901 RE B NM B
SEMANTIC-LEVEL 05 07
GOV S-PATTERN 07 1003 ARFS P AJFS P NMFS B
SEMANTIC-LEVEL 07 06

GOVERNOR-SEARCH
ANAPHORIC-ANALYSIS
GOV 02 EST 0502 NMMS B VVO 05
S-PAT 02 SM 02 SM B VD B AJMS B
DEP 02 GARCUN NMMS HU 11
GOV 02 BEAU AJMS
S-PAT 02 1103 ARMS P AJMS P NMMS B NR P
DEP 01 LF ARMS
DEP 03 QUI S'R 14
GOV 09 BEAU AJMS
GOV 01 LE ARMS
GOV 03 QUI S'R
S-PAT 03 1401 NR B VD B
DEP 04 VA VVO 06
GOV 04 VA VVO 06
S-PAT 04 0601 VD B RE B
DEP 05 A RE 09
GOV 05 A RE 09
S-PAT 05 0901 RE B NM B
DEP 07 ECOLE NMFS LD 10
GOV 07 ECOLE NMFS LD 10
S-PAT 07 1003 ARFS P AJFS P NMFS B
DEP 06 L' ARFS
GOV 06 L' ARFS

```

TEST SENTENCE # LF BEAU GARÇON! AIME LA BEAU FILLE #

LEXICON SEARCH

01	LF		ARMS				
02	BEAU		AJMS				
03	GARÇON		EMMS	HU		11	
04	AIME		VVD			02	
05	LA		ARFS				
06	BEAU		AJMS				
07	FILLE		HMFS	HU		10	

COMBINATION OF INFO

	LF		ARMS				
	BEAU		AJMS				
	GARÇON		EMMS	HU		11	
	AIME		VVD			02	
	LA		ARFS				
	BEAU		AJMS				
	FILLE		HMFS	HU		10	

FOR A STRING SUCH AS

01	LF		ARMS				
02	BEAU		AJMS				
03	GARÇON		EMMS	HU		11	
04	AIME		VVD			02	
05	LA		ARFS				
06	BEAU		AJMS				
07	FILLE		HMFS	HU		10	

MAIN PREDICATE VERB

DEPENDENT-SEARCH

04	0202	NM	B VD	B NM	B		
04	03						
04	07						
03	1103	ARMS	P AJMS	P NMMS	B NR	P	
03	02						
03	01						
07	1003	ARFS	P AJFS	P NMFS	B		
07	05						

GOVERNOR-SEARCH

GIVE UP PARSING

06

TEST SENTENCE # LF BEAU GARÇON! AIME LA BELLE FILLE #

LEXICON SEARCH

01	LF		ARMS				
02	BEAU		AJMS				
03	GARÇON		EMMS	HU		11	
04	AIME		VVD			02	
05	LA		ARFS				
06	BELLE		AJFS				
07	FILLE		EMFS	HU		10	

COMBINATION OF INFO

	LF		ARMS				
	BEAU		AJMS				
	GARÇON		EMMS	HU		11	
	AIME		VVD			02	
	LA		ARFS				
	BELLE		AJFS				
	FILLE		EMFS	HU		10	

FOR A STRING SUCH AS

01	LF		ARMS				
02	BEAU		AJMS				
03	GARÇON		EMMS	HU		11	
04	AIME		VVD			02	
05	LA		ARFS				
06	BELLE		AJFS				
07	FILLE		EMFS	HU		10	

MAIN PREDICATE VERB

DEPENDENT-SEARCH

04	0207	NM	B VD	B NM	B		
04	03						
04	07						
03	1103	ARMS	P AJMS	P NMMS	B NR	P	
03	02						
03	01						
07	1003	ARFS	P AJFS	P NMFS	B		
07	05						

GOVERNOR-SEARCH

ALPHABETIC-ANALYSIS

04	AIME		VVD				
04	0207	NM	B VD	B NM	02	B	
05	03	03 SH			07 SM		
05	GARÇON		EMMS	HU		11	
07	FILLE		EMMS	HU		10	
03	GARÇON		EMMS	HU		11	
1103	ARMS	P AJMS	P NMMS	B NR	P		
01	LF		ARMS				
02	BEAU		AJMS				
07	FILLE		EMFS	HU		10	
1003	ARFS	P AJFS	P NMFS	B			
05	03 SH						
05	LA		ARFS				
01	06		AJFS				
05	LF		ARMS				
02	BEAU		AJMS				
05	LA		ARFS				
06	BELLE		AJFS				