

SGLR-plusによる話者の対象認識構造を抽出する 英語文パーザの試作

高草木 伸高 宮崎 正弘

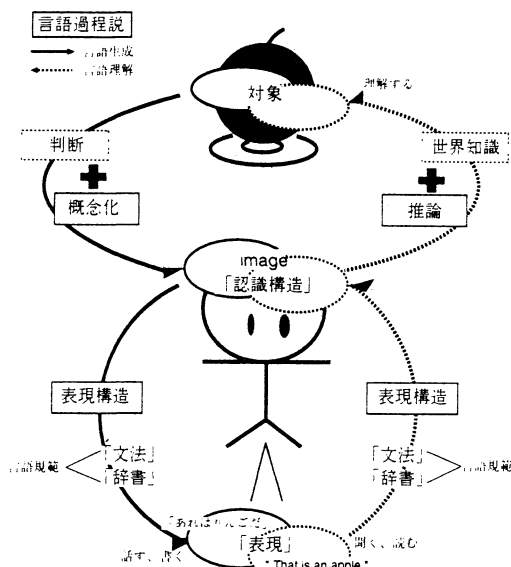
新潟大学大学院自然科学研究科

1 はじめに

計算機による自然言語処理において、より高度な自然言語の意味理解の実現には、自然言語の本質についての考察が必要である。そのため、人間の言語活動における「対象 ⇄ 認識 ⇄ 表現」という過程的構造に着目した時枝の言語過程説を発展的に継承した三浦の説 [1]、三浦説に基づき英語文法を構築した宮下の考え [2] に基づき、言語の統語構造に相当する表現構造と認識された対象世界を把握する認識構造を抽出する英語文パーザを試作した。一般化 LR 法 (富田法) [3] を Prolog 上に実現した拡張型の SGLR パーザ [4] である SGLR-plus [5] 上に、英語文法を補強項付きの DCG 形式で記述することにより、英語文から表現構造とともに、認識構造を抽出する英語文パーザのプロトタイプを試作した。

2 言語過程的説

人間の言語活動には言語生成と言語理解の2つがある。言語生成の場合、表現しようとする対象の客観的な姿を概念として表す。かつ、その対象に対し主観的で様々な判断を重ね合わせる。これにより人間は頭の中に認識構造を形成する。そしてその認識構造から発話、記述するなりして表現しようとするが、そこには必ず社会的な約束である言語規範 (文法、辞書、...) で拘束される。このようにして人間は言語を表現する。また逆に、ある言語表現を言語規範に基づいて聞いたり読んだりして、認識構造を頭の中に形成する。そしてその認識構造から主体の持つ世界知識と重ね合わせ対象の姿を推論することによって、言語表現は理解される。この様子を図 1 に示す。言語の過程的構造



3 認識構造構築のながれ

人間の主な言語表現は文字や発話によって表される。それらは文として存在し、表現と呼ばれる。それが各言語において文法上どのような構造か考えたものが表現構造である。その構造は「対象を客観的に捉えたものを概念として表す客体的表現」と「それに対する人間の主観的な判断、感情、客体認識の中の関係等を表す主観的表現」にわかれる。この分割作業は人間が言語規範によって聞いたり読んだりする過程で行なわれ、それを本能的に噛み砕いて「客体認識」と「主体判断」に変換する。そして認識構造を形成する。その様子を図 2 に示す。

A Experimental Parser for Extracting
Speaker's Recognition Structure
Nobutaka Takakusaki, Masahiro Miyazaki
Niigata University

例：You should not have asked him to do that.

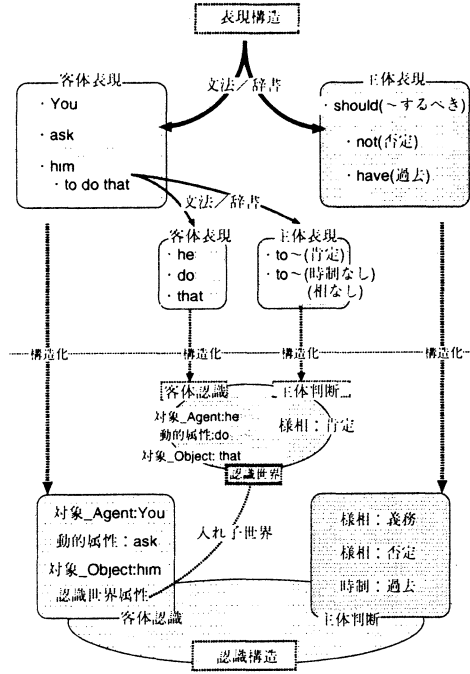


図2：表現構造から認識構造へのながれ

4 認識構造

認識構造は、対象を概念とその属性の集合で捉え関係づけたものを示す「客体認識」と、客体認識に対しての主観的で様々な判断、感情を含む情報である「主體判断」を重ね合わせて形成される。次にその要素について述べる。

客体認識を構成する要素

- ・ 動的属性 … 実体の動作を表す
- ・ 静的属性 … 対象の状態・性質を表す
- ・ 対象__Agent … 話者が注目している、また動作を行う対象の実体
- ・ 対象__Object … 動的属性の効果が及ぼす目標
- ・ 推移__Agent … 動的属性の効果が及ぼした後の対象 (__Agent) の姿
- ・ 受益者 … 動的属性の効果が及ぼした後の対象__Object の移動先
- ・ 時間/場所属性 … 対象__Agent, または各属性の存在する空間/時間情報

・ 認識世界属性 … 関係代名詞節/ where～節/ that～節など複文より考えられる、入れ子で存在する世界

主體判断を構成する要素

- ・ 様相 (modality) … 客体認識に対する表現主体の直接的な判断
- ・ 時制 (tense) … 表現主体と対象との時間的関係の認識を示す主體表現
- ・ 相 (aspect) … 実体の持つ動的属性が完了しているか、未完了か、進行中かなどを表す表現¹
- ・ 態 (voice) … 実体間における動的属性の波及の仕方を表した表現²
- ・ 焦点 (focus) … 表現主体が対象の、どの観点に興味/注目があるかを表す主體表現

5 SGLR-plus パーザ

パーザ本体としては SGLR (A Sequential Generalized LR Parser; 逐次型一般化 LR) パーザを用いる。この SGLR パーザは、横型探索をベースとした富田法を Prolog 上に記述したもので、先読み情報を利用することにより無駄のない高速な統語解析を実現する。また今回、関係代名詞節における痕跡の追跡、正規表現の記述、グラフ構造化などが可能となった SGLR の拡張版である SGLR-plus を使用した。その SGLR-plus の様子を図3に示す。

またパーザで用いる文法は、補強 CFG (Context-Free Grammar; 文脈自由文法) を Prolog 上で実現するために DCG (Definite Clause Grammar; 確定節文法) 形式で記述する。この DCG 形式で記述された文法は最終的にトランスレータによって Prolog プログラムに変換される。Prolog プログラムに変換して実行することにより、Prolog の基本計算機構をそのまま利用した自然言語処理を行なうことができる。DCG の文法形式は次のような特徴がある [3]。

- ・ 非終端記号を述語とみなし、それに任意この引数を持たせることができる。
- ・ 記号 "→" の右側の任意の場所に DCG の補強項と呼ばれる中括弧 '{ }' で括られた任意のゴール文を挿入することができる。

¹三浦によれば、相 (aspect) / 態 (voice) は、客体表現とあるが、ここでは、認識構造での見易さを考慮し主體判断に含める

²同上

- 記号"ー>"の右には、終端記号と非終端記号とが混在した任意の長さの記号列の記述ができる。つまり、CFGの一般形が記述できる。
- 補強項の中に、非終端記号に対応する述語の中で使われている引数を使うことができるので、統語的な制約が容易に記述できる。またそれにより、文法規則の精度をあげ統語解析速度をあげることも可能となってくる。

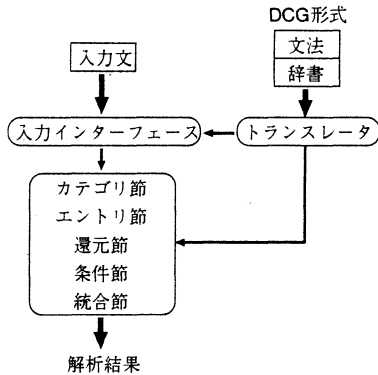


図3: SGLR-plus パーザ

6 文法

また今回使用した SGLR-plus の痕跡追跡機能により 40 の文法規則が省略可能となった。ここでは英語文を大きく 2 分する客体的主部と客体的述部の中核をなす"名詞句"と"動詞句"の文法規則の基本適用パターンについて図 4 に示す。

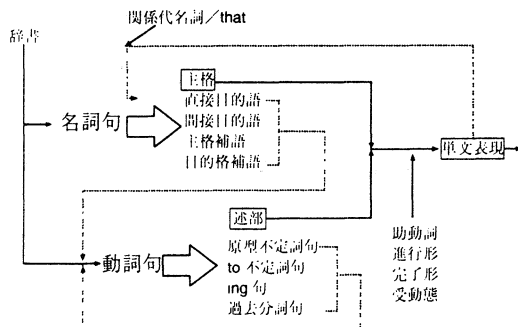


図4: 名詞句、動詞句の文法規則の基本適用パターン

7 辞書

今回は EDR 英語辞書を辞書引きプログラムに組み込んだことにより、英和辞典に通常の持っている単語を含む文の解析が可能となった。また、豊富な統語情報を用いることで統語的な曖昧性の解消、絞り込みが可能となった。全ての情報を、常に引数として渡し続けるわけにはいかないので、語形変化情報/文法情報/句型情報/に分割し、別々のファイルに DCG 形式で書き出し、必要時のみ補強項から呼び出すという手法をとる。また解析必要最低限の情報に関してはそのままパーザに渡すことにする。その辞書引きの様子を図 4 に示す。

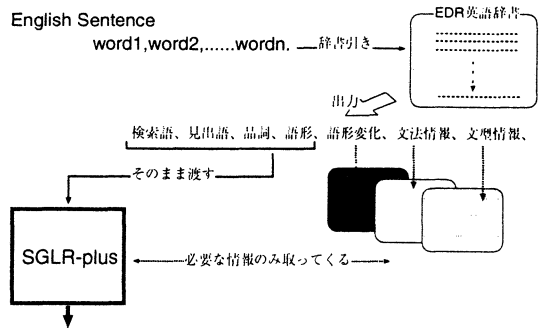


図4: 辞書引きの流れ

8 解析例

「You should not lend the book "that you borrowed"」("～"は関係代名詞節)を例として、表現構造(図4)と認識構造(図5)を抽出した。表現構造については、関係代名詞節内を先行詞と同じ構造が来るべき場を捜す痕跡機能によって、<\$gap>のように表示される。認識構造については、対象の概念として"The boy", 動的属性として"lend"が客体認識として存在する。そこへ主体判断である「様相は義務」「内容の否定」「時制は現在」「相は基準」が重ねられる。次に関係代名詞節に対する先行詞"the book"が持つ対象認識世界についてだが、"you"と"borrowed"と"the book"が客体認識として存在する。そして先行詞である"the book"があるべき"対象_Object"の場所には、痕跡機能によって先行詞である"対象_Object"と同じ構造を出力し、全体の構造をわかりやすいものへ近付けた。また<把握>は、話者が対象をどのような観点で概念化したかを表したものである。

```

|- 平叙表現 1
  |- 主格 1
  |   |- 人称代名詞
  |   |   |- EP1 -- < you >
  |- 助動詞
  |   |- EAV -- < should >
  |- 主体判断辞
  |   |- 否定判断辞_not
  |   |   |- 普通副詞
  |   |   |   |- ED5 -- < not >
  |- 述部 0
  |   |- 動詞句 how_
  |   |   |- 動詞
  |   |   |   |- EVE -- < lend >
  |   |   |- 直接目的語 1
  |   |   |   |- 名詞句 1
  |   |   |   |   |- 直接目的語 0
  |   |   |   |   |   |- 名詞句 0
  |   |   |   |   |   |   |- 冠詞
  |   |   |   |   |   |   |   |- EAR -- < the >
  |   |   |   |   |   |   |   |   |- 普通名詞
  |   |   |   |   |   |   |   |   |   |- EN1 -- < book >
  |   |   |   |   |- 関係代名詞
  |   |   |   |   |   |- EP5 -- < that >
  |   |   |   |- 平叙表現 1/直接目的語 0
  |   |   |   |   |- 主格 1
  |   |   |   |   |   |- 人称代名詞
  |   |   |   |   |   |   |- EP1 -- < you >
  |   |   |   |   |   |   |   |- 肯定判断辞_emp -- < $nil >
  |   |   |   |   |- 述部 0
  |   |   |   |   |   |- 動詞句 0
  |   |   |   |   |   |   |- 動詞
  |   |   |   |   |   |   |   |- EVE -- < borrowed >
  |   |   |   |   |- 直接目的語 0 -- $gap

```

図4 : 「You should not lend the book that you borrowed」
の表現構造

```

|- 認識世界
  |- 客体認識
  |   |- 動的属性 -- < lend >
  |   |   |- 対象_Agent
  |   |   |   |- 概念 -- < you >
  |   |   |- 対象_Object
  |   |   |   |- 把握 -- < the >
  |   |   |   |   |- 概念 -- < book >
  |   |   |- 認識世界属性
  |   |   |   |- 客体認識
  |   |   |   |   |- 動的属性 -- < borrow >
  |   |   |   |   |   |- 対象_Agent
  |   |   |   |   |   |   |- 概念 -- < you >
  |   |   |   |   |   |- 対象_Object
  |   |   |   |   |   |   |- 把握 -- < the >
  |   |   |   |   |   |   |   |- 概念 -- < book >
  |   |   |- 主体判断
  |   |   |   |- 様相 -- < 肯定 >
  |   |   |   |   |- 時制 -- < 過去 >
  |   |   |   |   |   |- 相 -- < 基準 >
  |- 主体判断
  |   |- 様相 -- < 義務 >
  |   |   |- 様相 -- < 否定 >
  |   |   |   |- 時制 -- < 現在 >
  |   |   |   |   |- 相 -- < 基準 >

```

図5 : 「You should not lend the book that you borrowed」
の認識構造

9 おわりに

本稿では、計算機による自然言語処理において、言語表現の持つ本質的な意味理解の実現を目指し、人間の言語活動である「対象 ⇄ 認識 ⇄ 表現」という過程の構造に着目した時枝誠記による言語過程説を発展的に継承した三浦つとむの説に基づいて、自然言語の本質について検討した。その検討に基づき SGLR-plus パーザを用い、表現構造と認識構造を抽出する英語文パーザのプロトタイプを試作した。今回 EDR 英語辞書を辞書引きプログラムとして利用することで、英和辞典に通常のっている単語を含む単文・複文の解析が可能と、基本動詞の細やかな文型束縛情報など、豊富な統語情報によって曖昧性解消にも有効となったとともに認識構造の抽出も可能となった。今後、重文への応用、多言語間での機械翻訳に有用のものへと検討する必要がある。また文法上でのいくつかの曖昧性解消のため意味解析との融合も必要となってくる。

10 謝辞

「EDR 英語辞書」の使用を許可された日本電子化辞書研究所の関係各位に感謝いたします。

参考文献

- [1] 三浦つとむ : 日本語とはどういう言語か, 講談社学術文庫 (1976)
- [2] 宮下真二 : 英語とはどういう言語か, 季節社 (1985)
- [3] 田中穂積 : “自然言語解析の基礎,” 産業図書, 1993.
- [4] 沼崎, 田中 : SGLR: 逐次型一般化 LR パーザの Prolog による実現, 情報処理学会論文誌, Vol.32, No.3, pp.396-403 (1991)
- [5] 五百川, 宮崎 : 痕跡処理のための逐次型一般化 LR パーザ SGLR の拡張, 言語処理学会第 4 回年次大会発表論文, pp.314-317 (1998)
- [6] 中島太郎 : 英語の仕組み, 日本実業出版社 (1994)
- [7] 大西泰斗/Paul Cris MacVay : ネイティブスピーカーの単語力 (1999)
- [8] 大西泰斗/Paul Cris MacVay : ネイティブスピーカーの英語感覚 (1999)