

非文法的な文章における意味を考慮した解析手法について

2Q-3

永松健司 田中英彦
 東京大学 工学部

1 はじめに

自然言語インタフェースや、現実使用される文章を処理する場合には、システムの規定する辞書や文法に合致しない、不適格な文章が問題になることがある。しかしながら、これまでの自然言語処理研究では文法に適合する文章のみを対象としてきたため、非文法的な文章が入力された場合には、解析に失敗して処理自体が行なわれなくなることがほとんどであった。

これに対して、いくつかの非文法的な文章の解析方法が提案されているが、その多くは構文解析における情報を利用したものであり、その文の意味までも考慮した処理を行なうものは少ない。

そこで今回、非文法的な文章を、その断片的な意味情報を利用して解析するシステムを試作したので報告する。

2 不適格文を含む文の解析の方法

このシステムの目的は、不適格な文(システムの持つ辞書/文法に合致しない文)に対しても、その文の部分部分を解析した意味情報から、全体の文の意味を解析することである。そのためには、構文解析に失敗しても、その時点でそれ以前の部分に対しても意味解析がなされていなければならない。また、一度、構文解析に失敗しても、その後の文章もまた引き続き解析する必要がある。そのために、構文解析に並行して意味解析を行ない、かつ全体の処理は意味解析部が主導して行なうこととしている(図1)。

文の解析においては、対話モデルに基づき、あらかじめ予想した文の格フレームを参照して、その必要な格を埋めていくように構文解析部が駆動される。また、構文解析部においても適宜、意味解析部を呼びだして意味情報を生成しつつ解析を進める。よって、構文解析の途中で失敗しても、その時点までで得られている意味情報を意味解析部は受けとることができる。また、意味解析部

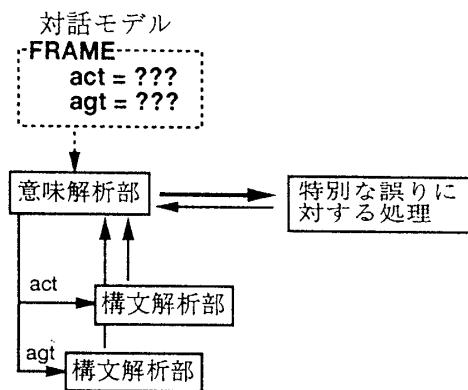


図1: 解析の流れ

では、格助詞等の構文情報とともに、意味制約も用いて柔軟に解析が行なわれる。これによって、不適格な文章に対して、たとえ構文解析において失敗し場合でさえも、最終的には何らかの意味情報が生成できることになる。

ここで、あらかじめ予想する文章には最初、複数候補を挙げておき、文の解析が進み、格関係が明らかになるにつれて、その候補を絞っていくようにする。この候補を絞っていく過程において、判断の根拠とするのは、名詞句ごとの意味素性および格助詞等の表記によってであるが、格助詞による格の指定よりも意味素性による制約を重視する。ただし、意味制約のチェックは極力厳しく行なうこととして、意味制約の違反が起こった場合などは別の処理を行なう。これは、意味制約のチェックを緩めてしまい意味的に間違った文までも通してしまうことを避け、意味の不整合を検出した場合には、不整合の原因を判断して適切な処理(比喩に対する処理、間違いを指摘する処理など)を行なえるようにするためである。

上述の解析方法によって、ある程度、不適格な文章を扱うことができるようになるのであるが、意味制約の違反などで、どうしても解析できない場合が起こりうる。また、この解析によって得られる意味情報は必ずしも正しいことは保証できない。よって、いくつかの特徴的な不適格性を持つ文章に対しては、各々記述したルールに基づいてそれを処理するようにする。これによって、人間にとって無意味な結果が出てくることを極力抑えるようにしている。

On parsing of ill-formed sentences that takes account of semantic informations

Kenji Nagamatsu Hidehiko Tanaka

Department of Electrical Engineering,

Faculty of Engineering, University of Tokyo,

7-3-1 Hongo, Bunkyo-Ku, Tokyo, 113, Japan

入力文章中に現れる不適格性の種類としては様々なものがありうるが、今回のシステムでは以下のようなものを例外的に扱うことにする。

未知語の検出：システムの持つ辞書にない単語が含まれた場合

意味制約の違反：名詞句の意味素性が、対応する格情報の持つ意味制約に違反している場合

以上の不適格性の例は、上述のような拡張された解析方法でも扱えないような場合に相当する。そして、これらに対しては、以下のようなルールを定め、解析の途中でこのような不適格性を検出した際に、対応するルールが駆動されて別途処理が行なわれる。

未知語の検出：未知語部分を抽出した後、述語の格情報および格助詞に基づいて、品詞および意味素性をできるかぎり推定する。

意味制約の違反：意味制約の違反をした語については、通常、そこに現れうる正しい意味素性の語句を文脈等から推定し、意味制約に違反した語との関係を調べることで、両者を修飾された名詞句とすることで、意味制約違反を解消する。

3 システムの動作

次に、実際にこのシステムが動作する様子を説明する。このシステムは、辞書としてEDR電子化辞書を用い、Prolog上にインプリメントしてある。

以下の例では、書店における店員との会話を対話モデルとして設定している。そこでの対話の流れとしては、「予約(予約する)」、「配達(届ける)」というものを想定する。そして、これらの動詞は以下のような格を取るものとする。

動詞	必須格(意味素性: デフォルト)
予約する	対象(人工物: 本)
届ける	対象(人工物: 本), 終点(場所: ???)

ここで、ユーザからの入力として、

夏目漱石の全集、東大、届けてください。(1)

という助詞が欠落している文が入力された場合、まず上の2通りの文が予想され、次に、その予想に基づき、対象(具体物)か終点(場所)等を解析しようとする。文章1の場合、意味制約のチェックを行なうことで、「夏目漱石の全集」が対象、東大が終点を示すものとして解析に成功し、東大=終点が判定できた時点で、予想される文は「配達(届ける)」に絞られ、動詞の解析が終了した時点でその予想が確認される。

次に、未知語を含む場合を見てみる。

夏目漱石の全集を注文してください。(2)

というような文が入力された場合、仮に「注文する」という動詞が未知であったとする。このとき、名詞句「全集」が格助詞および意味素性から対象格として解析されるが、終点としての場所を指定する名詞句がないので、「注文する」という動詞は、対象格を取る、「予約」の意味を持つ動詞であると見なされる。これは、予想される文が2つという単純化された場合だから、決定できたことであり、通常、この文からだけでは、意味の推定はできない。しかし、文章2の前に一文あって、

本を予約したい。
夏目漱石の全集を注文してください。(3)

であった場合には、第1文で予約する物を完全には確定していない状態にあり、第2文では前文の未確定部分を補う形になっていることから、動詞「注文する」は、「予約する」と同義の動詞として判断される。

最後に、意味制約を違反する語句を含む文章として、

夏目漱石を予約したい。(4)

という入力を与えられたとする。この場合、「予約する」の取る、対象格(人工物)に対して、解析して得られる対象格「夏目漱石」は意味素性: 人間(≠人工物)であり、意味制約の違反が生じる。

このとき、「予約する」の対象格となりうる名詞として「本」が推定される(これは、この対話モデルで「予約する」の対象格のデフォルトとして予め指定されている)。また、「本」には著者(人間)という属性を持つという知識を利用して、「夏目漱石を」は「夏目漱石の本を」を表す意味構造に変換される。

4 おわりに

不適格文に対して、その部分部分の意味情報を基にして、文全体を解析するシステムを作成し、その報告を行なった。このシステムでは、対話モデルからトップダウンに次の文の予想を行ない、意味解析部の主導のもとに格関係の解析を行なうことで、格助詞が欠落したり必須格が省略された文に対しても、その意味解析が可能となる。また、未知語や、意味制約の違反など、格関係からの推定では解析が困難な不適格性については、それらを処理するルールを別途設け、不適格性に応じて適切な処理を行なうことで人間にとって無意味な結果が出ることを極力抑えるようにしている。

参考文献

- [1] 松本裕治. 頑健な自然言語処理の現状と動向. 人工知能学会全国大会, pp. 75-78, 1993.
- [2] 豊田順一中村孝. 非文法的な日本語文を取り扱う意味主導型解析メカニズム. 情報処理学会自然言語処理研究報告, 1989. 70-8.