

英会話教育用知的C A Iシステムにおける
構文解析

3S-3

大里真理子+ 甲斐郷子+ 山本秀樹+ 椎野努+
沢山ゆかり++ 加藤正明++
+ 沖電気工業(株) ++ (株) 沖テクノシステムズラボラトリ

1. はじめに

英会話教育用知的C A Iシステムは、自然言語を使った自由な対話をシミュレートすることによって、学生の会話力を高めることを目標としている¹⁾。本稿では、上記の目標を達成するために構文解析に求められる機能とその実現手法について述べる。

2. システムの構成

図1にシステムの構成を示す。構文解析部は入力文の構文解析と意味表現への変換を行う。会話制御部は、その意味表現を使って次に発話する内容を決定し、出力文生成部ではそれによって英文を生成し出力する。

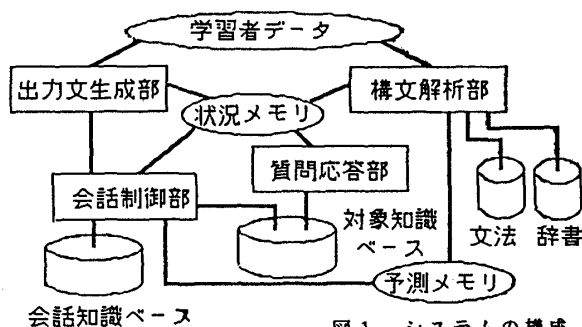


図1. システムの構成

3. 構文解析に要求される機能

本システムにおける構文解析部には、以下の機能が要求される。

(1) 高速性

会話のシミュレーションを用いて語学教育を行うには、システムからの応答が高速でなくてはならない。この応答の即時性によってシミュレーションは実際の対話により近くなり、学生の当意即妙性に対する教育的効果が高まる。そこで、システムの応答時間を短くするため、構文解析部においても処理の高速性が要求される。

(2) 誤りの同定機能

知的C A Iにおいては、学生の理解状況を正しく把握して、それを教育に反映することが必要である²⁾。そのために構文解析部では、学生の入力文の英文法的な誤りを同定する機能が必要である。ここで同定された誤りは、学生の理解状況を知るための情報の一つとなり、学生の能力に応じて会話のレベルを変更したり、誤りに対してメッセージを出力したりといった制御に用いられる。

predicate (動作)	: 動作	主対象	体系
agent (行為者)	: 動作の主体	対象	定まる
object (対象)	: 動作の対象	動作	ある
truth value (真偽)	: 真偽	動作	か
want (要求)	: 要求	動作	か
from (始点)	: 始点	動作	か
to (終点)	: 終点	動作	か
tense (時制)	: 時制	動作	か
method (方法)	: 方法	動作	か
place (場所)	: 場所	動作	か
others (細部情報)	: 細部情報	動作	か
sentence type (文の種類)	: 文の種類	動作	か

図2. イベントの構造

(3) 意味表現を生成する機能

構文解析部は、文法規則を使って文から解析木を生成し、それをもとに意味表現を生成する。この意味表現は、イベントと呼ばれる。イベントは格の概念の拡張であり、動作を中心とする知識表現である。内部構造として図2に示す12のスロットを持っている。各スロットは、それぞれ動作、行為者、対象などの意味を持ち、スロットとその値の対で固有の意味を表わしている。

次に、入力文から得られる意味表現の例を図3に示す。教材作成者は、教育の目的や会話制御部での処理に関連して、ある入力文からどのような意味表現を生成するかを決定する。従って、教材作成者の意図にもとづいた意味表現を自由に生成できなければならない。

また、会話の訓練を行うには、学生にスムーズな会話の流れを経験させる必要がある。そのため、誤りを含んだ文が入力されるたびに、意味生成に失敗して対話が中断してはならない。対話を継続させるためには、多少の誤りを含んだ文の入力に対しても、学生の本意を予想して正しい意味表現を生成することが必要となる。

(4) 綴り誤り訂正機能

学生が、綴りの打ち間違いや記憶違いなどで、正しくない綴りを入力することがある。そこで、それらの綴り誤りを訂正して解析する機能が必要となる。

次に、これらの機能を実現する手法について述べる。

4. 高速化

構文解析により木構造を作るアルゴリズムは現在までに数多く提案されている。ここでは、既存の構文解析アルゴリズムの中から処理速度の点で有効な Tomita's algorithm³⁾ をベースにし、さらに処理時間を短縮するために以下のことを行った。

(1) 候補木の削減

文を解析した結果多数の候補木ができる原因としては、文自体の係り受けの曖昧性、多品詞語により生じる曖昧性、取り扱う構文の数が増えることにより生じる曖昧性、がある。

処理の高速化のためには、これらの曖昧性をできるだけ取り除き、構文解析中に生成される候補木の中から不必要なもの(正しい意味を出さないもの)を、早い時点で削除することが

A Parsing Technique in ICAI System for English Conversation
Mariko OSATO+ Kyoko KAI+ Hideki YAMANOTO+ Tsutomu SHIINO+
Yukari SAWAYAMA++ Masaaki KATO++
+OKI Elec. Ind. Co., Ltd. ++OKI Techno Systems Laboratory Inc.

入力文	<u>Teach me the way to fill the form.</u>															
意味表現	event	<table border="0"> <tr><td>predicate</td><td>: fill</td></tr> <tr><td>agent</td><td>: student</td></tr> <tr><td>object</td><td>: ((form))</td></tr> <tr><td>tv</td><td>: t</td></tr> <tr><td>tense</td><td>: present</td></tr> <tr><td>method</td><td>: x</td></tr> <tr><td>s-type</td><td>: wh</td></tr> </table>	predicate	: fill	agent	: student	object	: ((form))	tv	: t	tense	: present	method	: x	s-type	: wh
predicate	: fill															
agent	: student															
object	: ((form))															
tv	: t															
tense	: present															
method	: x															
s-type	: wh															
入力文	<u>stay at this hotel.</u> <u>stay this hotel.</u> (誤り)															
意味表現	event	<table border="0"> <tr><td>predicate</td><td>: stay</td></tr> <tr><td>agent</td><td>: student</td></tr> <tr><td>object</td><td>: nil</td></tr> <tr><td>tv</td><td>: t</td></tr> <tr><td>tense</td><td>: present</td></tr> <tr><td>place</td><td>: hotel</td></tr> </table>	predicate	: stay	agent	: student	object	: nil	tv	: t	tense	: present	place	: hotel		
predicate	: stay															
agent	: student															
object	: nil															
tv	: t															
tense	: present															
place	: hotel															

図3. 意味表現の例

有効である。そこで、ここでは教育の対象とする学生のレベルを想定し、それに合わせて取り扱う構文を限定している。それにより句構造規則の数を押えることができる。

また、話題として取り扱う世界を限定することにより、各句構造規則の適用条件を厳しくすることができるので、係り受けの曖昧性を限定して処理することができる。また各文法の適用条件を制限すれば、それだけその文法における意味処理も制限して行い易くなり、結果的に処理速度の向上を図ることができる。また、多品詞語については、辞書を作る段階で、限定された世界で使われる品詞のみを選んで採録することにより、曖昧性を減らすことができる。

適切なレベルの教育を行うために、上記の制限範囲は、学生のレベルに応じて変更する。

(2) 熟語処理

熟語を含んでいる文を解析する場合、熟語を構成する単語毎に処理を行うと、多数の候補木が生成され、処理を遅らせる原因となる。

例えば、

in front of [前置詞 名詞 前置詞]

hand in hand [名詞 前置詞 名詞]

などを含む文をそのまま構文解析すると、多くの解析木が生成される。ここでは、in front of を前置詞に、hand in hand を副詞にとるといったように、連続した単語を一つの品詞として認識する処理、及び熟語とみなされた単語の並びに品詞、その他の辞書に記述されている情報を与え従来の解析過程に組み込む処理を行う。これにより、無駄な候補木の生成を防止し、処理の高速化を実現している。

5. 誤りの同定機能

入力文の英文法的な誤りを大別すると以下の2つがある。

① 語順以外に関する誤り

これには、冠詞の用法、前置詞の用法、時制、目的語、意味的な制約条件に関する誤り、等が含まれる。

ex. I have a apple.

ex. I belong in the basketball club.

ex. You go there yesterday.

ex. I bring _____.

ex. The room talked to me.

② 語順に関する誤り

ex. I want a room good.

ここでは、上記のそれぞれの誤りの同定法について述べる。

もともと単語は、この動詞は目的語を必要とする、あるいは、この動詞の主語は人間に限られる、などの語彙素性を持っている。①の誤りは、単語の語彙素性によって誤りであることがわかる。そこで構文解析部で使用する単語辞書中に、単語の語彙素性を表わすセマンティック・マーカを記述する。構文解析中にこのセマンティック・マーカを参照する機能を持たせることにより、①の誤りは同定できる。②の誤りについては、誤りの語順を含んだ文法規則を本来の文法規則に加える。この誤りを含んだ規則を部分木を持つ解析木ができれば、その文は語順に関して誤った文であると同定できる。

6. 意味表現生成処理

2. (2)で述べたように、意味表現生成時には、教材作成者の意図に基づいた意味表現を生成することと、誤りを含む入力文から正しい意味表現を生成することが必要である。

入力文から教材作成者の意図する意味表現を生成するには、単語の文中での位置や単語の持つ語彙素性の組み合わせを用いることが必要である。また、4. で述べた英文法的な誤りのうち①は、単語の持つ語彙素性の組み合わせで表わすことができる。

そこで、意味生成時に単語のセマンティック・マーカを参照する機能を持たせる。これにより、単語の持つ語彙素性の情報を参照しながら意味を生成することができる。

また、4. で述べた誤りのうち②については、誤りの語順を含んだ文法規則の意味生成時に、正しい意味表現を生成する処理を行う。

7. 綴り誤り訂正機能

従来、綴り誤りを訂正するには、主として統計的なデータを用いる方法がとられてきた。しかし、この方法では、統計的データに頼っているため、1文字誤りでも、もとの正しい意図された単語を捨てる可能性がある。そこで、1文字の誤り（1文字の置換、挿入、欠落、隣どうしの文字の入れ替え）については、統計的データを必要とせず、アルゴリズム的に、綴り誤りを訂正する方法をとった。この方法を用いるならば、意図した単語が辞書中に含まれている時には必ず、見つけることができる。また、2文字以上の誤りについては、統計的データを用いる方法で誤りの訂正を行っている。

8. 終りに

本稿では、知的CAIシステムの構文解析に要求される高速性、誤りの同定機能、意味表現生成機能、綴り誤り訂正機能を実現する手法について述べた。以上の手法は、ELIS上で開発中の英会話教育用知的CAIシステムの構文解析部に実現されている。

参考文献

- 1) 大里、他: 語学会話教育用ICAIシステムの構成. 情報処理学会第33回全国大会講演論文集 (1986).
- 2) 大槻: 知的CAIの現状と問題点について, 人工知能学会第1回全国大会チュートリアル講演 (1987).
- 3) Tomita, M: Efficient Parsing for Natural Language, Kluwer Academic Publishers (1985).