キーボード会話文の構文解析文法について

4F-6

野上宏康　吉村裕美子　熊野明　天野真家
(株)　東芝　総合研究所

1．はじめに

科学技術文献・マニュアルなどに現れる文に対する構文解析文法は盛んに研究され，既に商品化されている機械翻訳システムに組込まれている．今後は，加速度的な国際化の進展に伴い，自動翻訳電話等に応用可能な会話文の解析の研究が重要になってくると考えられる．

本稿ではその第一ステップとして，英語のキーボード会話文の構文解析文法について述べる．

2．キーボード会話の環境

キーボード会話とは，会話者がキーボードとディスプレイを用いて会話を進めるものである．会話者が自分の発話をキーボードから入力すると，その発話は自分のディスプレイと相手のディスプレイに表示され，履歴として残るようになっている．

本稿では，このような環境で入力される会話文を対象として考察する．

3．キーボード会話文の構文解析文法

ここでは，キーボード会話に現れる現象とその構文解析文法について，文書の文に対する構文解析文法と比較しながら述べる．

3－1．キーボード入力に起因する現象

キーボード会話では入力の省力化として，短縮形，小文字化，アポストロフィの省略等が頻繁に用いられる．その例を図1に示す．

短縮形，小文字化に対しては，辞書に小文字化したものを登録することにより，アポストロフィの省略に対しては形態素解析で処理することにより対処可能である．したがって，これらの現象に対して解析文法を新たに記述する必要はない．

短縮形
　pls(please)，u(you)，cn(can)
小文字化
　i，japan，tokyo，jal
アポストロフィの省略
　cant，wont，dont

図1　　キーボード入力に起因する現象の例

3－2．文の断片による表現

キーボード会話では，文の断片のみ，または文や断片がカンマで区切られて連続する現象がよく出現する．その例を図2に示す．

このような現象は，文書の文に対する解析文法では一般に解析できない．これらは，図4に示す文法を記述することにより解析が可能になる．ここで，開始記号としてUを用いている．但し，この文法は従来の文法より優先度を低くして適用する必要がある．それは，英語には多品詞語が多いことなどから，誤った解析を行う可能性があるからである．

・married，2 children，many friends.
・Very good，where did you look it up?
・Yes，but when I was only 12 years old.

図2　　文の断片による表現例

Grammar for parsing in conversation through keyboard
Hiroyasu NOGAMI，Yumiko YOSHIMURA，Akira KUMANO，Shin-ya AMANO
TOSHIBA Corp.

S→NP・VP
VP→V ・NP
NP→N
・

図３　文書の文に対する文法

U→A
U→A・カンマ・U

A→S
A→VP
A→NP
・

図４　文の断片表現に対する文法

３－３．文の境界が明確でない表現

この表現は，ユーザが２文以上を一度に入力する場合に文の切れ目にピリオドを省略してしまうことにより現れる．図５にその例を示す．音声による会話文の解析を考えた場合，ピリオドは発声されないから，このことからもこの現象に対処することは重要である．

このような現象を従来の解析文法で解析できないのは，従来の文法は１文の範囲が認識できていることが前提となっているからである．この現象の解析は，図６に示す文法を記述することにより可能になる．但し，この文法も従来の文法よりも優先度を下げて適用する必要がある．これに対し，解析前に１文を予想して切り出し，解析する方法も考えられるが，この方法では切り出す位置の検出が非常に困難である．

That is OK it is beautiful to look at.

図５　文の境界が明確でない表現

U→S・U
U→S

図６　文の境界が明確でない表現の文法記述

３－４．非文やタイプミス等の現象

キーボード会話の場合には，文書の文に比べて非文やタイプミスが多くなる．その例を図７に示すが，この種の文は一般に解析が不可能である．しかし，会話をできるかぎり自然に続けることは重要である．また全体の解析に失敗しても，句単位で解析に成功していれば会話の文脈中では理解可能なことが非常に多い．したがって，このような現象に対処しておくことは重要である．

図７の第１文，第２文のような現象に対しては，U→A・U，U→Aという文法を記述することにより，句単位の解析には成功する．第３の文のような現象の解析に失敗するのは，top-down・left-to-right の解析の場合，"／" より右側には到達できないからである．このような文に対しては，A→＊（＊は任意の単語を意味する）という文法を更に記述することにより句単位の解析が可能になる．この文法は，解析できない部分を飛び越える機能を果たす．但し，これらの文法は最も優先度を低くして適用する必要がある．

・Are married to 2 what?
・It means learn much in a very short time.
・Really/ I am an honest person.

図７　非文やタイプミスの例

４．おわりに

本稿ではキーボード会話文の構文解析文法について述べた．本文法を我々が開発した自動翻訳文字電話[1][2]に用いて実験し，有効に機能することが確認できた．

参考文献

[1] 天野，武田，長谷部，平川；「自動翻訳文字電話の構想」情報処理学会第36回全国大会予稿1U-2(pp.1215～1216)
[2] 旭岡，吉村，三池，野上；「自動翻訳文字電話の翻訳について」情報処理学会第36回全国大会予稿1U-4(pp.1219～1220)