

パラフレーズ機能を備えた 自然言語インタフェースシステム

酒井 桂一[†] 池田 裕治[†] 藤田 稔[†]

文書整形システム LaTeX のコマンド使用法を対象とする自然言語インタフェースシステムについて述べる。自然言語インタフェースにおいては、(1)システムが受け付ける文のカバーレージの狭さ、(2)ユーザの発話意図理解の難しさ、(3)ユーザとシステムの間が生じる誤解などの問題点がある。さらに、自然言語インタフェースの応用システムとしては、インタフェース部分が頑丈であること、ユーザごとに容易に調整できるように柔軟であることなどが求められる。これらの問題点に対処するために、まず、ユーザごとに独自に用いられる用語を含めた、特定のドメインで用いられる用語を4段階に階層化した。そして、各階層間の用語変換辞書を用いてパラフレーズすることによって、システムが受け付ける文のカバーレージを広げるアプローチをとった。また、(A)伝達モード解析と主動詞の分類に基づいて、ユーザの発話意図を抽出する機能、(B)ユーザの発話中に未知語や省略格が存在するなど、ユーザの発話がシステムにとって不完全な場合に不完全な箇所をユーザに問い返す機能を備えた。さらに、他のインタフェースとの融合を試みた。記号の一覧や書式例を求める質問に対する応答として、文書整形結果をディスプレイに表示する機能を設けた。

A Natural Language Interface with Paraphrasing

KEIICHI SAKAI,[†] YUJI IKEDA[†] and MINORU FUJITA[†]

In this paper, we present Japanese natural language interface system for the command usage of LaTeX: a document preparation system. Natural language interfaces have several problems: narrow coverage of the expressions they receive, difficulty of understanding user's intention, misunderstanding between a user and a system, and so on. And its applications require the robustness and the customization flexibility of the interface part. To solve these problems, we classified the terms used in a specific domain into 4 levels, including a level that each user can enter his own words. The system paraphrases words by using a dictionary in each level transit. Also it extracts user's intention based on modality analysis and category of main verb, through a follow-up question when the user's utterance includes unknown words or ellipses. To combine natural language interface with other applications, displaying symbols and format examples is used for replying as well as answering literally in natural language.

1. はじめに

我々は、文書整形システム LaTeX[®] のコマンド使用法を対象とした自然言語インタフェースの研究を行っている。

自然言語インタフェースの問題には、システムが受け付ける文のカバーレージの狭さ、ユーザの発話意図を正しく理解することの難しさ、ユーザが持つ知識とシステムが持つ知識のギャップによる誤解などがある。

自然言語インタフェースの応用システムが今後広く使われるためには、これらの技術的な問題を考慮することは勿論のこと、応用システム全体としてインタフ

ェース部分が頑丈であること、応用システムがユーザごとに容易に調整できるように柔軟であることが必要であろう。

従来の研究では、対話の流れを制御するために、ユーザモデルを用いた談話管理方式や、Grice の会話の公準¹⁾ に則って生成する応答文を調整する研究などが主であった。例えば UNIX Consultant²⁾ では、概念記述言語 (KODIAK) を用いて、対象である Unix Command の概念とその関係を記述する。そして、あらかじめランク分けされた知識やコマンドをユーザが知っているかどうかによりユーザレベルを推定する。また、スタックを用いて話題を制御する例³⁾ や、焦点に基づいて生成する説明文を変える例⁴⁾ などがある。

これらの研究では、対象に対するユーザの知識レベルを考慮することで、対話の流れを制御したり応答内

[†] キヤノン(株)情報システム研究所
Canon Inc. Information Systems Research Center

容を変えたりしているが、システムが受け付ける文のカバーレッジを広げるためのものではない。

それに対して、意味的なカバーレッジを広げることを目的として、標準的な文からの語彙的、構文的、意味的な距離を計算して、適切なアプリケーションを選択する Full Language Approach を用いている例⁹⁾もあるが、その選択をユーザごとに調整することは実現していない。

我々は、システムが受け付ける文のカバーレッジを広げ、かつ、ユーザごとに調整が可能な「パラフレーズ解析」を提案する。まず、パラフレーズを形態素レベル、係り受けレベル、意味表現レベルの三つに分け、形態素レベルのパラフレーズ解析部および、その辞書構築法を作成した¹⁵⁾。また、ユーザの発話意図を理解するために、日本語のモダリティ^{6),7)}を基にしたムード解析による発話意図抽出部を作成した¹⁴⁾。対話の流れの制御については、ユーザモデルを用いる方式を考慮するが、モデルに基づく強い制御は行わない方式をとった。

さらに、システム全体の使い勝手を良くするために、自然言語によるインタフェースと他のインタフェースとの融合の可能性についても考慮した。特に文書整形システムのコマンドを対象としていることから、コマンドのシミュレート機能を実現することが有用であろう。そのために文書整形結果をディスプレイに表示するアプローチをとった。

2. パラフレーズ解析

自然言語インタフェースでは、ユーザが自分の言葉で質問、応答などの入力を行えることが重要である。一方、システムの内部に保持される知識は、画一的な形式で、対象とするドメイン特有の用語を用いて記述される。そこで、ユーザによる多様な表現をシステム内部の画一的な表現に『言い換える』機能が重要となる。このような機能を「パラフレーズ解析」と呼ぶ。

2.1 パラフレーズ解析のレベル分け

パラフレーズ解析として、以下の三つのレベルを考える。なお、形態素レベルのパラフレーズについては、2.2節で詳しく述べる。

【形態素レベルのパラフレーズ】

同義語や異表記を登録した辞書を用

意し、形態素レベル、すなわち、単語対単語のパラフレーズを行う。

【係り受けレベルのパラフレーズ】

品詞や構文情報から、係り受けレベル、すなわち、句対句のパラフレーズを行う。

例えば、「文を引用する」と「文の引用を行う」のように、『サ変動詞』とその名詞形との間のパラフレーズや、「行の間のスペース」と「行間スペース」のように、『AのB』とその『複合名詞』間のパラフレーズなどが挙げられる。

【意味表現レベルのパラフレーズ】

深層格や意味素性などの情報から意味表現レベルのパラフレーズを行う。

例えば、「改行する」と「行を変える」のように格関係が異なるものや、「右端に移動する」と「右に詰める」のように異なる語に『端』の意味が含まれるもののパラフレーズが挙げられる。

2.2 形態素レベルのパラフレーズ解析の詳細

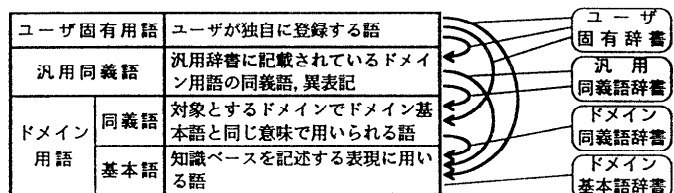
形態素レベルのパラフレーズを実現するに当たり、図1に示すように、ある特定のドメインで使われる用語を階層化した。さらに、用語の階層に基づいてパラフレーズを行うためのパラフレーズ用辞書を設定した。

2.2.1 用語の階層

【ドメイン用語】

ドメインを特定すると、汎用辞書には記載されていない語が用いられたり、汎用辞書の定義と異なる意味で用いられたりする。このように特定のドメインで用いられる語を「ドメイン用語」とする。

しかし、原語が外国語のものや解説書などが複数存在するものでは、解説書によって訳語が異なったり、原語をそのまま用いたり、あるいは原語を単にカタカナに置き換えた語を用いたりして、統一が取れていないものがある。



※図中の矢印(→)は各階層の用語がパラフレーズされる様子を示す

図1 特定のドメインにおける用語の階層とパラフレーズ用辞書
Fig. 1 Hierarchy of terms in the specific domain and dictionaries for paraphrase.

例えば, LaTeX の開発者である Lamport 著のマニュアル⁸⁾の 'type style' に対し, その訳本⁹⁾(以下, 単に『訳本』と呼ぶ)では「書体」という訳語を与えている. 一方, 『訳本』⁹⁾に先駆けて出版された解説書『楽々 LaTeX』¹⁰⁾では, 「文字フォント」という用語を与えている. また, 『TeX 入門』¹¹⁾では, 「字体」という用語を用いている.

そこで, 「ドメイン用語」のうち, 知識ベースを記述する表現に用いる語を「ドメイン基本語」とする. また, 「ドメイン基本語」と同じ意味で用いられる語を「ドメイン同義語」とする.

【汎用同義語】

ユーザは, 「ドメイン用語」を用いずに, 汎用辞書に記載されている「ドメイン用語」の同義語や異表記を用いることがある.

例えば, ドメイン用語である「アブストラクト」, 「要約」の同義語で汎用辞書から得られるものとしては,

概要, 抄録, 要旨, 梗概, あらすじ
などが挙げられる.

このように, 汎用辞書に記載されている「ドメイン用語」の同義語, 異表記を「汎用同義語」とする.

【ユーザ固有語】

各ユーザごとに記号の俗称や略称など, 「ドメイン用語」でも「汎用同義語」でもない単語を用いることがある. このように, 個々のユーザが自由に用いる語を「ユーザ固有語」とする.

2.2.2 パラフレーズ解析用辞書

【ドメイン基本語辞書】

知識ベースの記述に用いる「ドメイン基本語」(の見出しのみ)を登録する. 本システムでは, 『訳本』⁹⁾, 『楽々 LaTeX』¹⁰⁾, 『TeX 入門』¹¹⁾の索引から主に抽出した. また, 複合名詞の場合は, そのものと展開したものをすべて登録した. 内容は以下のとおりである.

- (a) 『訳本』⁹⁾の索引に存在する語…622 語
- (b) (a)に無く, 『楽々 LaTeX』¹⁰⁾『TeX 入門』¹¹⁾の索引に存在する語…135 語
(そのほとんどは, 『訳本』⁹⁾中で英単語として出現)
- (c) (a), (b)に無く, 『訳本』⁹⁾の本文の記述に現れる語…8 語

【例】 「外枠」, 「書式」など

【ドメイン同義語辞書】

『訳本』⁹⁾, 『楽々 LaTeX』¹⁰⁾, 『TeX 入門』¹¹⁾の索引から「ドメイン基本語」の同義語を「ドメイン同義語」として抽出し, 以下の書式で登録した.

〈辞書エントリ〉 ::= 〈左辺〉 → 〈右辺〉.
 〈左辺〉 ::= 〈ドメイン同義語〉 |
 〈ドメイン同義語〉, 〈左辺〉.
 〈右辺〉 ::= 〈ドメイン基本語〉.

名詞の場合, 「原語をカタカナ表記したもの」と「原語の訳語」との同義語が約 70% を占め, 動詞の場合には, いわゆる「サ変動詞」と「和語動詞」との同義語が約 60% を占めた.

【例】 要約 → アブストラクト
 文字フォント, 字体 → 書体
 作る → 作成する
 印刷する → 印字する

【汎用同義語辞書】

汎用辞書から「ドメイン基本語」あるいは「ドメイン同義語」の同義語, 異表記を「汎用同義語」として抽出し, 以下の書式で登録した.

〈辞書エントリ〉 ::= 〈左辺〉 → 〈右辺〉.
 〈左辺〉 ::= 〈汎用同義語〉 | 〈汎用同義語〉, 〈左辺〉.
 〈右辺〉 ::= 〈ドメイン基本語〉 | 〈ドメイン同義語〉.

【例】 概要, 要旨, あらすじ → 要約
 梗概, 抄録 → アブストラクト
 移す, 動かす → 移動する

【ユーザ固有辞書】

「ドメイン用語」あるいは「汎用同義語」の同義語としてユーザが個々に以下の書式で登録する.

〈辞書エントリ〉 ::= 〈左辺〉 → 〈右辺〉.
 〈左辺〉 ::= 〈ユーザ固有語〉 |
 〈ユーザ固有語〉, 〈左辺〉.
 〈右辺〉 ::= 〈ドメイン基本語〉 | 〈ドメイン同義語〉 |
 〈汎用同義語〉.

【例】 ニョロ → チルダ
 グニャ括弧 → ブレース

3. 応用システム

2.2 節で述べた形態素レベルのパラフレーズを実装した LaTeX コマンドヘルプシステムを NeXT コンピュータ上に作成した. 図 2 に本システムの構成を示す.

長期的な構想としては, 今後の拡張を意識して, 「談話管理」, 「プラン管理」あるいは「協調的応答生

成」を包含した「基本モデル」を考えている¹³⁾が、図2では、実現した部分のみを示す。

本システムでは、対話ウィンドウから入力された入力文を文解析した後、パラフレーズ解析を行う。対話管理部では、3.2節で述べるユーザの発話意図の抽出、3.3節で述べる、質問一応答などの「対話対」の入れ子に対処する対話対管理を行って、知識ベースからデータを抽出するか、ユーザに情報を求める『問い返し』を行うかを判定する。出力生成部では、問い返しの質問文、データ抽出結果の説明文あるいは書式例を生成し、対話ウィンドウあるいはプレビューアに出力する。なお、現在の質問応答機能は以下のとおりである。

- 操作内容からコマンド（環境）名を答える。
- 記号などの一覧をプレビューアに出力する。
- 書式例をプレビューアに出力する。

3.1 パラフレーズ解析部

入力文の各形態素に対し、「ユーザ固有辞書」、「汎用同義語辞書」、「ドメイン同義語辞書」、の順で各パラフレーズ用辞書を検索し、パラフレーズ用辞書に登録されている語であれば、置換を行う（図1）。その後、「ドメイン基本語辞書」を検索し、登録されていない形態素が存在する場合は、「未知語」とし、3.3節で示すように、『ユーザに問い返す』ことにより、「ドメイン基本語」を得る。

現在、パラフレーズ用辞書に登録されている語数は以下のとおりである。

ドメイン基本語辞書 764語

表1 伝達ムードの分類

Table 1 Classification of modality.

伝達ムード名	文末表現の例	事柄の主体	伝達ムードの主体
働き掛け	命令	(命令形), なさい	聞き手
	依頼	て下さい, てくれ	聞き手
	勧誘	(よ)う, ましょう	聞き手+話し手
表出	意志	(よ)う, ましょう	話し手
	希望1	たい	話し手
	希望2	てほしい	聞き手
	否定希望	てほしくない	聞き手
問い掛け	依頼	て下さるか, てくれるか て下さらないか, てくれないか	聞き手
	勧誘	(よ)うか, ましょうか	聞き手+話し手
	意向	(よ)うか, ましょうか	話し手
	希望1	たいか, たくないか	聞き手
	希望2	てほしいか, てほしくないか	話し手

ドメイン同義語辞書 129語

汎用同義語辞書 75語

3.2 発話意図抽出部

質問応答で用いられる対話には、単純な質問文、応答文だけでなく、依頼や希望などを表す表現（伝達ムード）を含む文が用いられる場合がある。そこで本システムでは、伝達ムードに基づいた発話意図の抽出を行っている。

3.2.1 伝達ムードの分類

伝達ムードについては、仁田が言語学の立場から『話し手の発話時における伝達の態度』として分類を行っている⁹⁾（以降、『仁田の分類』と呼ぶ）。ここでは、伝達ムードを働き掛け、表出、述べ立て、問い掛けの四つに大分類し、それぞれについて文法的特徴を述べている。そこでシステムに実装するために、これを参考にし、以下に示す点について考慮して、働き掛け、表出、問い掛けについての細分類を行った。

- (a) 『仁田の分類』⁹⁾での表層上の格格についての制約に対し、深層格の立場から事柄の主体と伝達ムードの主体に分離して考える。
- (b) 問い掛けは、働き掛けおよび表出に疑問の終助詞「か」が付与可能かどうか、可能な場合に事柄および伝達ムードの主体がどう変化するかを判定する。

細分類の結果を表1に示す。

(a)に関しては、例えば、「行って下さい。」という発話に対しては、『仁田の分類』⁹⁾で表層上の格格を「聞き手」に制約しているのに対し、「行く」という事

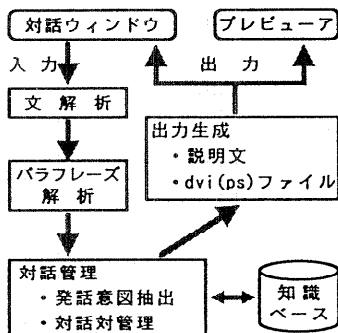


図2 LaTeX コマンドヘルプのシステム構成図
Fig. 2 System structure of LaTeX command help.

表 2 質問応答の対話における発話意図
Table 2 Intentions obtained from main verb and modality in the Q&A dialog.

伝達 ムード名	働き掛け		表 出			問 い 掛 け			
	命令	依頼	意思	希望 1	希望 2	依頼	意向	希望 1	希望 2
知覚動詞	提 示	提 示	承 認	要 求	提 示	提 示	要 求	提 示	要 求
教示動詞	要 求	要 求	承 認	提 示	要 求	要 求	提 示	提 示	提 示
調査動詞	要 求	要 求	承 認	要 求	要 求	要 求	提 示	提 示	提 示

柄の主体が「聞き手」であり、その事柄を伝達するムード、依頼の働き掛けの主体が「話し手」であることを示す（記号“⇒”はムード解析の結果を示す）。

行って下さい⇒ムード：依頼（主：話し手）
事 柄：行く（主：聞き手）

また、『仁田の分類』⁶⁾で一つにまとめられていた希望の表出を事柄の主体の違いから希望 1 と希望 2 に細分類した^{*}。

行きたい⇒ムード：希望 1（主：話し手）
事 柄：行く（主：話し手）

行ってほしい⇒ムード：希望 2（主：話し手）
事 柄：行く（主：聞き手）

(b)に関しては、働き掛け、表出の分類のうち、『仁田の分類』⁶⁾の命令、禁止、依頼的禁止には終助詞「か」をつけることはできない。それ以外のものに関して、終助詞「か」をつけることによって、伝達ムードおよび事柄の主体がどのように変化するかを調べると、以下になることがわかった。

依頼、勧誘、意向（意志の問い掛け）については、伝達ムードの主体、事柄の主体とも変化しない。

行ってくれるか⇒ムード：依頼（主：話し手）
事 柄：行く（主：聞き手）

希望 1、希望 2 については、伝達ムードの主体、事柄の主体とも「聞き手」と「話し手」が入れ替わる。

行きたいか⇒ムード：希望 1（主：聞き手）
事 柄：行く（主：聞き手）

行ってほしいか⇒ムード：依頼（主：聞き手）
事 柄：行く（主：話し手）

3.2.2 発話意図

質問応答においてなされる発話のうち、3.2.1 項で分類した伝達ムードを伴うものの事柄の主動詞として、以下に示す 3 種類に大別した。

知覚動詞 「知る」、「理解する」、「見る」など、『主体が情報を得る行為』を表すもの

教示動詞 「教える」、「示す」、「見せる」など、『主体が相手に情報を与える行為』を表すもの

調査動詞 「調べる」、「求める」など、『主体が情報を明らかにする行為』を表すもの

3.2.1 項で示した伝達ムードの分類と上記の 3 種類の動詞の組み合わせによって決定する発話意図の一覧を表 2 に示す。

表 2 では、命令と依頼の働き掛け、希望 2 の表出、依頼の問い掛けは、上記 3 種類の動詞に関して同一の発話意図であるとしている^{*}。

また、希望 1 の表出と希望 2 の表出とを比較すると、知覚動詞および教示動詞が希望 1 の伝達ムードを伴う場合には、その発話意図は、それぞれ情報の提示と要求になる。それに対し、希望 2 の伝達ムードを伴う場合には、発話意図が逆転し、それぞれ要求と提示になる（記号“⇔”は発話意図が逆転することを示す）。

知りたい（要求）⇔知ってほしい（提示）

教えたい（提示）⇔教えてほしい（要求）

一方、調査動詞は、希望 1、希望 2 のどちらの伝達ムードを伴ってもその発話意図はともに情報の要求となる（記号“=”は同一の発話意図を示す）。

調べたい（要求）=調べてほしい（要求）

また、意志の表出では、発話意図を承認としているが、これは質問応答の文脈では、要求の意図を持つ発話に対して承認の意図を持つ発話に用いられることを示す。

教えましょう（承認）

上記のもの以外に、いわゆる疑問の名詞述語文（「AはBか。」）の発話意図を質問とする。また、肯定的名詞述語文（「AはBだ。」）の発話意図を応答とする^{**}。

* これらのそれぞれ（例えば命令と依頼）の間には、いわゆる「発話の力の強さ¹²⁾」に差があると考えられるが、本稿では考慮しない。

** これは、3.3 節で述べる「問い返し」の質問に対する応答としてのみ受理し、それ以外の状態でユーザがこのような発話を行った場合には、「処理不能です」というメッセージを返す。

* 希望には、例えば「彼に来てほしい」のように、事柄（彼が来る）の主体が第三者（彼）になるものがあるが、これについては本稿では考慮しない。

3.3 対話対管理部

本システムでは、指示照応、省略などの談話処理の前段階として、対話対管理を行っている。「対話対」とは、挨拶—挨拶、質問—応答、提示—選択などの関係を持つ発話の対のことであり、対話の内部構造は、基本的にこの「対話対」によって構成されると考えられる³⁾。

質問応答システムでは、ユーザの質問に対し、システムが何の障害もなく応答できる場合には、一問一答として対話対が完成する。しかし、システムが質問を受け取り、応答を生成するまでの過程で、何らかの障害（例えばユーザの質問中にシステムにとって情報が不十分な場合）が発生した場合には、応答できないので対話対が完成しない。このような場合、システムがユーザに「問い返し」の質問を行い、ユーザから障害を解消する応答が得ることができれば、システムはユーザの元の質問に対する応答を行う。

本システムは以下の場合に「問い返し」を行う。

- (1) パラフレーズ用辞書に存在しない「未知語」が入力文中に現れた場合
- (2) 必須格要素が省略されており、かつその推定が不可能な場合
- (3) 検索するデータの「内容」が欠落している場合

以上のような問い返しの処理を可能とするために、図3(a)に示すようなユーザとシステムの発話意図を格納するスタック（以後、対話対スタックと呼ぶ）を用いることにした。

まず、発話抽出の結果、ユーザの発話意図が質問（要求）となった場合には、その質問をユーザのスタックにプッシュする。その質問に対し、システムが応答可能な場合には、システムの発話意図を応答（提示）とし、システムのスタックにそれをプッシュすることに

より、質問—応答の対話対が完成するので、両者をポップし、対話対スタックは空になる（図3(b)）。

一方、ユーザの質問に対し、システムに何らかの障害が発生し、応答不可能な場合には、その障害を解消すべく、システムがユーザに「問い返し」の質問を行い、システムのスタックにその質問をプッシュする。

システムの「問い返し」の質問に対し、ユーザが応答を与えた場合には、対話対が完成する。従って、両者をポップし、元の質問に復帰する。ユーザの応答が適切で、元の質問の障害が解消された場合には、システムは元の質問に対する応答を行うことができ、両者をポップすることによって対話対スタックが空になる（図3(c)）。

3.4 対話例

本システムにおける対話例を以下に示す。

(1) 通常の一問一答の場合

Usr: 脚注を出力するコマンドを知りたい。

Sys: footnote コマンドを使って下さい。

Usr: フォントをゴシック体に変えるコマンドは gothic ですか。

Sys: いいえ、gt コマンドです。

Usr: フォントを明朝体に変えるコマンドは mc コマンドですか。

Sys: はい、そうです。

(2) 未知語が現れた場合には、『[未知語]』とは、どういう意味ですか。』という質問を問い返す。

Usr: テンテンをつけるコマンドを教えてください。

Sys: 『テンテン』とは、どういう意味ですか。

Usr: 傍点のことです。

Sys: '!' コマンドを使って下さい。

(3) 必須格要素が省略された場合には、省略された必須格の入力を促す。

Usr: 移動するコマンドを知りたい。

Sys: 何を移動するのですか。

Usr: 文字列です。

Sys: どこに移動するのですか。

Usr: 右端です。

Sys: flushright 環境か raggedleft 宣言を使って下さい。

(4) 検索するデータの内容が欠落している場合には、「どんな（検索タイプ）ですか」という質問を問い返す。

Usr: コマンドを知りたい。

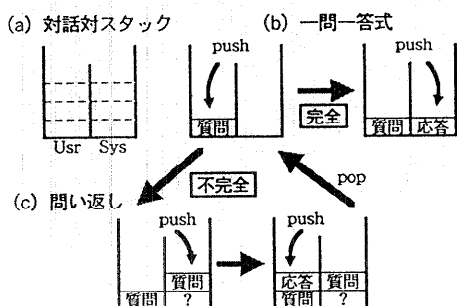


図3 対話対スタックの動作

Fig. 3 Act of utterance-pairs stack.

Sys: どんなコマンドですか.

Usr: 下線を引くコマンドです.

Sys: underline コマンドを使って下さい.

(5) 記号などの一覧や書式例はプレビューアに出力する (図 4 参照).

Usr: 矢印の一覧を見せて下さい.

Sys: プレビューアに出力します.

Usr: itemize 環境の書式を見たい.

Sys: プレビューアに出力します.

図 4 に, NeXT コンピュータ上での LaTeX コマンドヘルプシステムの処理例を示す. なお, 現在, 登録されているコマンド (環境) 数は以下のとおりである.

操作内容⇒コマンド (環境) 名	111 コマンド
記号などの一覧 (プレビューア出力)	308 コマンド
書式例 (プレビューア出力)	20 コマンド

4. 評価および考察

作成した応用システムを評価するために, 以下に示

す実験を行った.

15 名の LaTeX ユーザを被験者として選び, 与えたコマンド名を失念した場合に, 被験者がどのように質問するかを収集し, パラフレーズと発話意図に分けて考察する.

4.1 パラフレーズ解析の評価

5 個の LaTeX コマンドに対し, 被験者の質問文からコマンドの操作内容の部分を取り出し, 知識ベースで記述している表現にするのに必要なパラフレーズの種類と回数を集計した. 集計結果を表 3 に示す.

a. 有効回答数

被験者がそのコマンドを知らなかったり, 別のコマンドと誤解して答えたりしたものを除いた件数.

b. パラフレーズなし

パラフレーズを行わなくても知識ベースの表現と一致した件数.

c. 形態素レベル

形態素レベルのパラフレーズを行った表現の数. その後, 係り受けレベルおよび意味表現レベルのパラフレーズが必要なものも含む. また, 表中で字下げされ

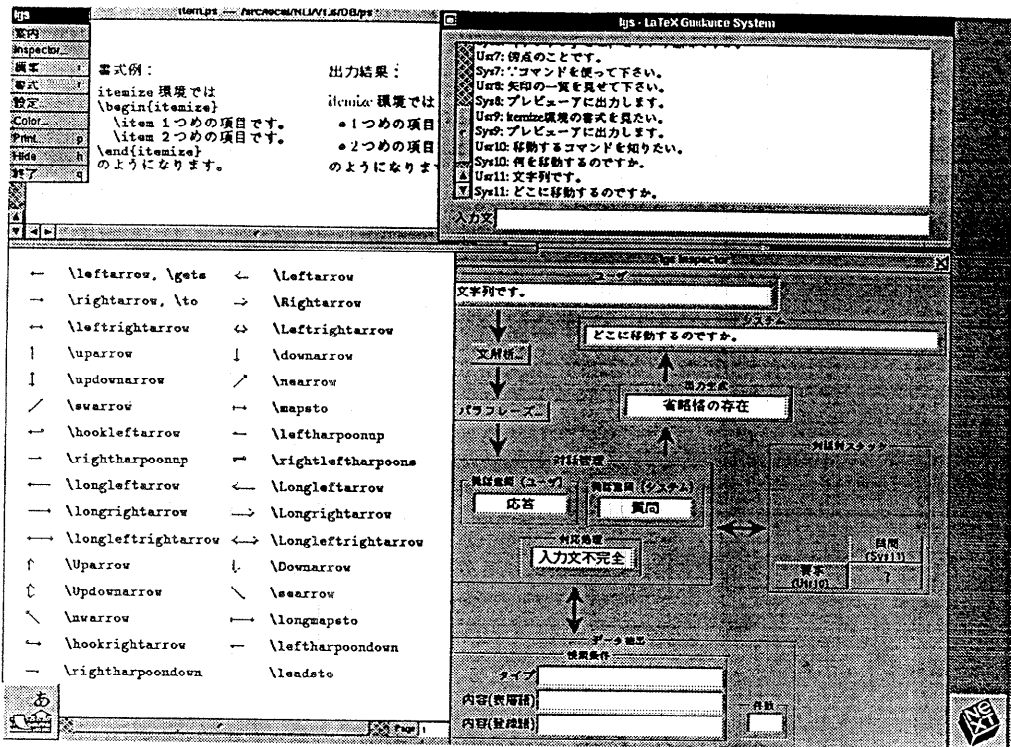


図 4 LaTeX コマンドヘルプシステムの処理例
 Fig. 4 Hard copy of LaTeX command help system with displaying symbols and a format example.

表 3 パラフレーズ解析の評価結果
Table 3 The evaluation of paraphrasing.

	(1)	(2)	(3)	(4)	(5)	合計
有効回答数	14	14	15	13	12	68
パラフレーズなし	2	4	3	7	7	23
形態素レベル	8	8	5	5	2	28
ドメイン同義語	10	8	7	4	2	29
汎用同義語	6	0	1	2	0	8
形態素レベルまで	6	11	7	10	7	43
係り受けレベル	2	0	4	0	0	6
意味表現レベル	5	2	2	2	2	13
意味表現レベルまで	11	13	11	12	9	56

(1) bf, (2) footnote, (3) enumerate, (4) flushleft, (5) quotation

たドメイン同義語および汎用同義語の項目の値は、「ドメイン同義語辞書」および「汎用同義語辞書」でパラフレーズを行った形態素の「のべ件数」である。

d. 形態素レベルまで

形態素レベルのパラフレーズまでで知識ベースの表現と一致した件数。

e. 係り受けレベル

係り受けレベルのパラフレーズを行われる件数を人手で集計した。

f. 意味表現レベル

係り受けレベルのパラフレーズを行われる件数を人手で集計した。

g. 意味表現レベルまで

意味表現レベルのパラフレーズまで行くと知識ベースの表現と一致する件数。

表 4 より、形態素レベルおよび全レベルのパラフレーズを行うか否かによるデータ抽出の成功率は以下ようになる。

パラフレーズなし (b/a) …34%

形態素レベルのパラフレーズ (d/a) …63%

全レベルのパラフレーズ (g/a) …82%

パラフレーズを行わない場合には、ユーザが用いた表現の約 1/3 しかコマンド名が得られないのに対し、形態素レベルのパラフレーズを行うことにより、その倍近くの表現からコマンド名を得ることができた。従って、パラフレーズ解析がシステムが受理する文のカバーレッジを広げる手段として、有効といえるであろう。

なお、パラフレーズを行っても知識ベースの表現と一致しないものとして、以下のものがあつた。

(1) ドメイン用語として異なる語を用いた場合

例えば、footnote に対し、「脚注」の代わりに「コメント」を用いた表現があつたが、両者が異なるドメイン用語であつたため、パラフレーズされず、未知語にもならなかつた。

(2) 他のコマンドと比較する表現を用いた場合

例えば、enumerate (番号付きの箇条書き) に対し、「itemize の黒丸の代わりに番号を付ける」という表現があつたが、他のコマンドと比較する表現には対処していない。

(3) 知識ベースに登録していない表現を用いた場合

例えば、quotation (段落引用) に対し、「テキストの両側を字下げする」という表現があつたが、このような表現が知識ベースに登録されていなかった。

4.2 発話意図抽出の評価

「情報の要求」の意図で発話されたもののうち、約 80% については、3.2 節で述べた方法で抽出できた。

発話意図を抽出できなかった例を以下に示す。

(1) 「対事ムード」に基づくもの

対事ムードとは、「話し手の事柄に対する把握の仕方」を表す表現で、推量、伝聞、良好などがあるとされているが、現在のところ、システムに実装するには整理が不十分である。

【例】「××するにはどうしたら良いか。」

「どうしたら××できるか。」

(2) 「間接発話行為」によるもの

直接的な表現でなく、その発話の裏で「情報の要求」を行っていると考えられるものである。

【例】「××できない。」

(⇒だから、できる方法を知りたい。)

「××するコマンドを知っているか。」

(⇒知っていたら教えてほしい。)

5. おわりに

本稿では、ユーザのさまざまな発話に対応できる自然言語インタフェースを目標として試作した、文書整形システム LaTeX のコマンドヘルプについて述べた。

知識ベースに登録されている表現にユーザのさまざまな表現を一致させるために、形態素レベル、係り受けレベル、意味表現レベルの三つのレベルに分けてパラフレーズ (言い換え) を行うことを提案し、形態素レベルのパラフレーズをシステムに実装した。また、実際の質問応答の結果を評価することによって、パラフレーズ機能が有効であることを確認した。

謝辞 本研究を行う機会を与えてくださいました知能工学研究部田村秀行部長に厚く御礼申し上げます。また、日頃御討議いただく知能工学 13 研究室の皆様感謝いたします。また、質問応答例の収集に協力していただいた社内の皆様に感謝します。

参 考 文 献

- 1) Grice, H. P.: *Logic and Conversation, Syntax and Semantics*, Vol. 3, pp. 41-58, Academic Press (1975).
- 2) Wilenski, R. et al.: The Berkley UNIX Consultant Project, *Computational Linguistics*, Vol. 14, No. 4, pp. 35-84 (1988).
- 3) 宮地ほか: 話題管理機能を持つ対話システムの試作, 情報処理学会知識工学と人工知能研究会, AI-38-7 (1985).
- 4) 垣内ほか: 焦点の概念を導入したマニュアルレスシステムの説明機能について, *Proceedings of the Logic Programming Conference '86*, pp. 19-26 (1986).
- 5) 渡辺ほか: 複数のアプリケーションと応答可能な自然言語インターフェース・システム, 情報処理学会自然言語処理研究会, NL-65-3 (1988).
- 6) 仁田義雄: 現代日本語文のモダリティの体系と構造, 日本語のモダリティ, pp. 1-56, くろしお出版 (1989).
- 7) 益岡隆志: モダリティの構造と疑問・否定のスコープ, 日本語のモダリティ, pp. 193-210, くろしお出版 (1989).
- 8) Lamport, L.: *LaTeX: A Document Preparation System*, p. 242, Addison-Wesley Publishing Co. (1986).
- 9) Cooke, E., 倉沢良一監訳: 文書処理システム LaTeX, p. 241, アスキー出版局 (1990).
- 10) 野寺隆志: 楽々 LaTeX, p. 247, 共立出版 (1990).
- 11) 大野義夫: TeX 入門, p. 260, 共立出版 (1989).
- 12) 久米ほか: 日本語対話文における発話意図の解法の方法, 第 36 回情報処理学会全国大会論文集 3 T-9 (1988).
- 13) 池田ほか: ガイダンス機能を備えた自然言語インタフェースシステム (1)―概要―, 第 42 回情報処理学会全国大会論文集, 4 C-5 (1991).
- 14) 酒井ほか: ガイダンス機能を備えた自然言語

インタフェースシステム (2)―発話意図抽出処理―第 42 回情報処理学会全国大会論文集, 4 C-6 (1991).

- 15) 酒井ほか: ガイダンス機能を備えた自然言語インタフェースシステム (3)―パラフレーズ解析―第 43 回情報処理学会全国大会論文集, 3 G-8 (1991).

(平成 4 年 1 月 27 日受付)
(平成 5 年 4 月 8 日採録)



酒井 桂一 (正会員)

1963 年生. 1987 年京都大学工学部電気工学第二学科卒業. 1989 年同大学院修士課程修了. 同年キャノン(株)入社. 現在, キャノン(株)情報システム研究所勤務. 自然言語処理

の研究に従事.



池田 裕治 (正会員)

1960 年生. 1982 年東京農工大学工学部数理情報卒業. 1984 年同大学院修士課程修了. 同年キャノン(株)入社. 1986 年京都大学工学部電気工学第二教室受託研究員. 現在キャノ

ン(株)情報システム研究所勤務. 文字認識, 人工知能, 自然言語処理の研究に従事. 人工知能学会, ソフトウェア科学会, 日本認知科学会各会員.



藤田 稔 (正会員)

1952 年生. 1975 年東京工業大学工学部制御工学科卒業. 1977 年大学院システム科学専攻修士課程修了.

同年富士通(株)入社. 1983 年キャノン(株)入社. 以来, 自然言語処理の

研究に従事. 現在, 情報システム研究所知能工学研究部に勤務. 計測自動制御学会, 計量国語学会, 国語学会各会員.