

日本語による建築材料に関する質問応答システム†

戸内 順 一††

日本語で、関係モデルのデータを操作しつつ、建築材料に関する質問応答を行うシステムについて述べている。従来、いくつかの自然言語によるデータベース操作システムが作成されているが、実用に耐えるものは少ない。また、多くは、英語によるもので、日本語によるものも少ない。そこで本研究でも、日本語によるデータベース操作システムを開発し、実用の可能性を追求した。

本システムは次のような特徴をもつ。すなわち、日本語で関係モデルデータの検索・追加・更新を行える。また、データベース操作のための中間言語として、単語辞書中の意味情報の係り受け構造をもとに、写像型言語を作成している。

具体的には、利用者が、材料の情報および材料の使用部位に関する質問を日本語で入力すると、データベースを検索して、その答えを返す。また、平叙文で、材料に関する情報を入力すると、データベースの更新・追加を行う。

本システムは、東京大学大型計算機センター HITAC 8800/8700 の HLISP を使って実動化した。

1. はじめに

計算機と対話をする言語として、フォートラン、コボルなどの形式言語ではなく、日常使用される自然言語を用いることは、計算機技術者や研究者の長年の夢であった。

近年、それらは、最近の多くの研究者や技術者の努力により、次第に実現されつつあるが、まだ、研究室レベルで止まっており実用困難な状況である。

というのは、一般に自然言語が正しく理解されるためには、言語が語られる環境や文脈まで認識する必要があり、それが、計算機にとって苦手な分野であるからである。

しかし、問題をデータベースに対する質問応答に限れば、対象となる世界や、使用される語も限られるので、計算機にも理解しやすい。

それで、自然言語でデータベースを操作するシステムがいくつか開発されている。英語を使ったものとしては、LUNAR¹⁾、REQUEST²⁾、TORUS³⁾、LADDER⁴⁾ などがある。また、日本語によるデータベース照会システムとしては、日本アイ・ビー・エムの「ヤチマタ」⁵⁾ がある。

データベースの普及にともない、エンドユーザの為のデータベース操作言語として、自然言語に近いものが、今後、増々要求されていくと考えられる。

そこで、本システムにおいても、自然言語によるデータベース操作システムを開発し、実用の可能性を追求した。

2. システムの概要

本システムは、日本語で、関係モデルのデータを操作しつつ、建築材料に関する質問応答を行うことを可能にしている。主に、建築材料の性能および建築材料の使用部位に関する質問に対して応答することができる。

またデータベースの検索だけでなく、データベースへのデータの追加や修正も可能であり、通常のデータベース操作言語のもつ基本的な機能を日本語で為すことができる。なお、ここでは、非常にシンプルな個人用データベースを仮定しており、同時実行制御やそれにとまなう排他制御・リカバリなどは考慮していない。

データベースは関係モデルで作成してある。関係モデルは、高度のデータ独立性をもち、簡潔で、専門外の人でも理解しやすく、種々の高水準のデータ操作言語を考えることができる。関係モデルでは、専門家でない利用者とデータベースとのインタフェースを重視するため、自然言語は有効な手段となりうる。また、データの独立性が高いことも自然言語には適している。コード自身も、理想的なデータベース操作言語を自然言語と考えると、RENDEZVOUS⁶⁾ を提案した。

自然言語解析において、英語は、語順により、文の意味が決定されるが、日本語では、語順は、あまり重

† A Japanese Question-Answering System on Building Materials by JUNICHI TOUCHI (Faculty of Science and Engineering, Tokyo University of Science).

†† 東京理科大学理工学部経営工学科

要でなく、文節ごとの係り受け構造が、意味の決定において重要となる。そこで、本システムでは、構文を句構造規則でなく、係り受け規則により解析している。すなわち、係り受け解析を行いつつ、対応する意味構造を生成成長させ、最終的に、文全体の意味を表わすと考えられるデータベース操作のための写像型コマンドを作成する。このとき、意味情報は、単語辞書の中に埋め込まれており、意味情報どうしの係り受けにより、意味構造を構成する属性名や属性値が決定される。LADDER 等のように、意味的拘束を直接統語規則に埋め込んだ意味文法の手法との類似点もあるが、LADDER が、語順にもとづく ATN パーサによる構文解析とパターンマッチングにより意味構造を作成しているのに比べ、本システムでは、単語辞書にある意味情報の係り受け構造をもとに意味構造を作成している。

最終的に作成される意味構造は写像型コマンドとなる。写像型コマンドは、関係にある属性からほかの属性への写像と考えて設計されており、まず、条件が与えられた属性にだけ着目して、表を探索し、与えられた属性値に写像する。関係論理型コマンドや関係代数型コマンドに比べて、量記号が不要であり、直観的に分かりやすいなどの長所をもつ。特に量記号をもたないため、自然言語からの変換が容易である。

3. データベース

本データベースは関係モデルで作成されている。

関係モデルはデータの独立・データ構造の簡潔さなどの特徴をもつ。自然言語でデータベースを操作するためには、特に、データの物理的な構造からの独立は重要な条件である。

本システムでは、材料性能と BE に関する 2 種類の関係表を用意している (表 1)。

BETBL は材料と、その材料が使用される場所の関係を示している。たとえば、塩化ビニールとゴムシートは床材料であり、ルーフィングは屋根材料である。

SEITBL は材料とその性能の関係を示している。たとえば、ゴムシートの耐摩耗性の等級は 8 であり、リノリウムシートの耐熱性の等級は 5 である。なお、等級は 0~9 の 10 段階にグレード分けしており、9 に近いほど、性能が良い事を示す。また、数字の前に ¥ 記号をつけた場合はあいまいな等級を表わす。たとえば、¥8 は「8 ぐらいの等級」という意味である。

表 1 関係表

Table 1 Relational table.

BETBL (ZAI BE)

ZAI	BE
塩化ビニール	床
ゴムシート	床
リノリウムシート	床
ゴムタイル	床
ルーフィング	屋根
ハードボード	天井

⋮

SEITBL (ZAI SEI TOU)

ZAI	SEI	TOU
ゴムシート	耐摩耗性	8
リノリウムシート	耐熱性	5
ゴムタイル	耐水性	8
アスファルトタイル	耐透水性	8
アスファルトタイル	耐アルカリ性	8

⋮

4. データベース操作コマンド

本システムでは中間言語のデータベース操作コマンドとして、SEL, TEST そして MAKE の 3 つの写像型言語を採用している。本言語は、SQL⁵⁾ と基本的に似ているが、自然言語の意味を反映しやすくように設計した。

SEL は SQL の SELECT に似ており、what 型疑問文に対応している。MAKE は SQL の INSERT と UPDATE を兼ねたような述語であり、追加、更新を行う。これは平叙文に対応する。TEST のような機能をもつ述語は SQL にはないが、YES/NO 疑問文に対応するコマンドとして設計した。

3 つのコマンドの仕様を図 1 に示す。条件の論理演算子として、AND と OR、そして関係演算子として、

```

SEL      関係名  属性名リスト  条件
MAKE     関係名  属性名リスト  属性値リスト  条件
TEST     条件
条件
(論理演算子 (関係演算子 1 属性名 1 属性値 1)
              (関係演算子 2 属性名 2 属性値 2)
              ⋮
            )

```

図 1 データベース操作コマンドの仕様

Fig. 1 Command for relational database.

EQ, NEQ, GE, LE などをもつ。

各コマンドの処理は次の通りである。

SEL コマンドは関係表より条件を満足する組をさがし、指定された属性の値を返す。

MAKE コマンドは関係表に条件を満足する組があれば、その組の属性の値を更新する。条件を満足する組がなければ、関係表に条件を満足する組を追加する。

TEST コマンドは、関係表に条件を満たす組があればTを返し、なければFを返す。

5. コマンドの生成過程

データベース操作コマンドを構成する属性は、単語辞書の意味情報をもとに決定される。意味情報には、単語の意味的分類を示す意味品詞と単語の意味そのものを示す意味単語がある。

使用部位 (BE) に関する単語や材料 (ZAI) に関する単語は、意味品詞、意味単語がそれぞれ属性名、属性値となる。一方、性能 (SEI) や等級 (TOU) に関する単語は、意味単語の係り受けにより属性名と属性値が決定される。これらの決定された属性名と属性値のうち、あるものを条件とし、あるものを更新や検索の対象とする。最後に、コマンドを構成する属性の種類によって関係名を決定し、コマンドを生成する。

次に、システムの制御の流れを追いながら、コマンドの生成過程を記述する。本システムは図2のような構成をしている。

1つの文が入力されると、まず語彙解析に制御が渡される。語彙解析は単語辞書を参照して、入力文を文節単位に切り分けるとともに、文節の文法的情報 (品詞、態) や意味情報を得る。単語辞書はバックトラッキングしながら最長一致でサーチされる。単語辞書は名詞、形容詞、動詞などからなる詞の部分と、助動詞、助詞、活用語尾などからなる辞の部分に分けられる。詞は次のような構成をしている。

(単語 品詞 接続情報 意味品詞 意味単語)

ここで

品詞 ADV (副詞) NN (名詞) ADJ (形容詞)
ADJV (形容動詞) VRB (動詞)

接続情報 ADJ (形容詞活用語尾)

ADJV (形容動詞活用語尾)

VB5 (5 段活用語尾)

VBSA (サ変活用語尾)

VBS1 (下一段活用語尾)

VBK1 (上一段活用語尾)

L2 (助詞) X (接続なし)

意味品詞 ZAI (材料) TOU (等級)

BE (使用部位) SEI (性能)

SEI1 (性能1) SEI2 (性能2)

TOU1 (等級1) TOU2 (等級2)

その他

接続情報は、辞に対する接続を表わしている。

単語辞書の例を図3に示す。たとえば、単語 ASPHALT-TILE は品詞が名詞で助詞に接続する。意味品詞は材料で、意味単語は ASPHALT-TILE である。また、単語 SOUTOU は品詞が副詞で、辞には接続

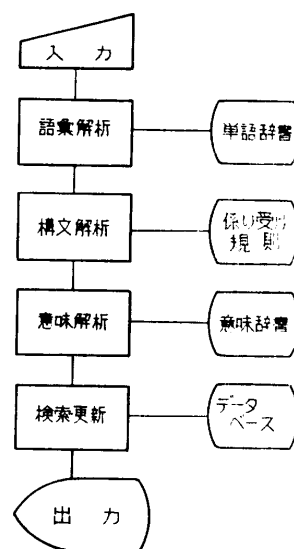


図2 システムの構成

Fig. 2 System construction.

(ASPHALT-TILE NN L2 ZAI ASPHALT-TILE)
(MOSAIC-TILE NN L2 ZAI MOSAIC-TILE)
(GUM-SHEET NN L2 ZAI GUM-SHEET)
(SOTOKABE NN L2 BE SOTOKABE)
(KAIKOUBU NN L2 BE KAIKOUBU)
(YAKAMASI ADJ ADJ SEI2 NOISY)
(YAWARAKA ADJ ADJ SEI2 SOFT)
(KIWAMETE ADV X TOU1 A)
(HIZYONI ADV X TOU1 B)
(SOUTOU ADV X TOU1 B)
(HUUATU NN L2 SEI1 KAZE)
(KETURO NN L2 SEI1 KETURO)
(KEMURI NN L2 SEI1 KEMURI)
(SEINOI NN L2 SEI NIL)
(KAZYUU NN L2 NIL NIL)
(MAMOU NN L2 SEI1 MAMOU)
(OUKI ADJ ADJ TOU2 H)
(TAE VRB VB5 SEI2 BEAR)

図3 単語辞書の例

Fig. 3 Word dictionary.

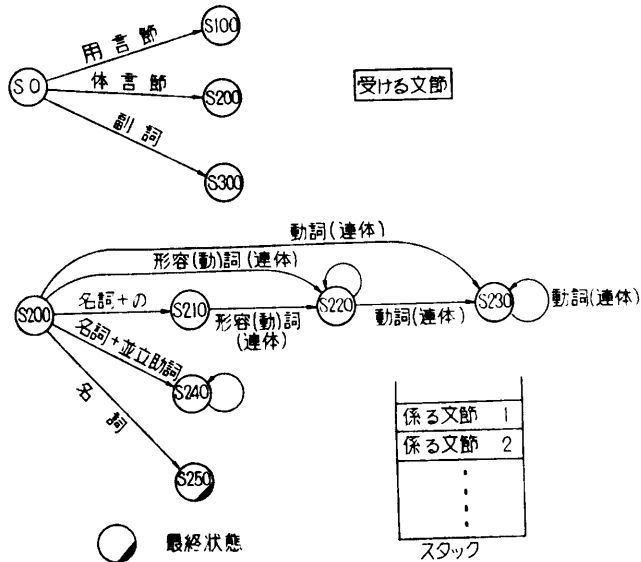


図 4 状態遷移図とスタック

Fig. 4 Transition network and stack.

しない。意味品詞は等級1で意味単語はBとなる。

構文解析では、語彙解析で得られた文法的情報をもとに、プッシュダウンスタックを使って節文どうしの係り受け木を作成する。係り受け規則は状態遷移図で表現されている(図4)。初期状態を S0 とし、受ける文節に対応して、S100, S200, S300 のいずれかに状態を移す。次にスタックのトップにある文節を取り出し、その文節が状態を移すための条件をみたせば、係り受け可能と判断し、状態を移動して、スタックのトップをはずして、係り受け木を作成する。以上の処理を、スタックが空になるか、最終状態または係り受け不能となるまで繰り返す、最後に、受ける文節をスタックに入れる(図5)。

構文解析で作成された係り受け木をもとに意味辞書を参照しながらデータベース操作コマンドを作成するのが意味解析である。意味辞書の形式は次のようである。

(意味単語1 意味単語2 属性名 属性値)

これは、意味単語1と意味単語2が係り受け関係にあるときの属性名と属性値を示している。意味辞書の例を図6に示す。意味品詞 SEI1 と SEI2 の係り受けにおいては、意味単語 MAMOU が意味単語 BEAR

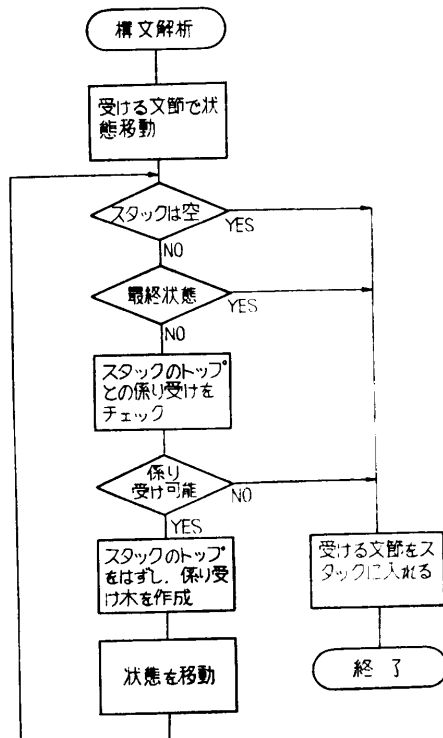


図 5 構文解析プログラム

Fig. 5 Syntax analysis program.

SEI1 と SEI2 の係り受けにより SEI (性能) が決定される場合

SEI1	SEI2	属性名	属性値	
(OTO BLOCK	SEI	ON 1)	防音性	
(OTO ABSORB	SEI	ON 2)	吸音性	
(MAGE BEAR	SEI	GA 5)	耐曲げ性	
(KAZE BEAR	SEI	GA 6)	風圧性	
(MAMOU BEAR	SEI	GA 4)	耐摩耗性	
(MIZU BEAR	SEI	MI 2)	耐水性	
(NETU CUT	SEI	NE 2)	断熱性	
(NETU BEAR	SEI	NE 1)	耐熱性	
(HI DEFEND	SEI	HI 1)	防火性	

TOU1 と TOU2 の係り受けにより TOU (等級) が決定される場合

TOU1	TOU2	属性名	属性値
(A H	TOU	¥ 9)	
(B H	TOU	¥ 8)	
(C H	TOU	¥ 7)	
(D H	TOU	¥ 6)	
(NIL M	TOU	¥ 5)	
(D L	TOU	¥ 4)	
(C L	TOU	¥ 3)	
(B L	TOU	¥ 2)	
(A L	TOU	¥ 1)	

図 6 意味辞書の例

Fig. 6 Semantic dictionary.

に係るとき、または、反対に BEAR が MAMOU に係るとき、属性名は SEI (性能)、属性値は、GA 4 (耐摩耗性) となる。もし、構文解析から渡された係り受けが意味辞書中に定義されていないならば、スタックに係り受け木を構成する係る文節と受ける文節をスタックに入れもどす。1つの係り受け木の処理が終わったなら、構文解析に制御を渡し、次の係り受け木の処理を行う。以上の処理を1つの文の解析が終わるまで繰り返す。1つの文の処理が終わった時点で、スタックに1だけ文節が残っていれば正常終了であり、それ以外は異常終了である。異常終了の場合は、意味不明の通知を利用者にして、再入力を求める。

構文解析と意味解析の関連で、コマンドが生成される過程を図7に示す。図7の①、②、③の左側は、構文解析で作られる係り受け木であり、右側は、係り受け木からつくられる意味構造である。④は、これらの意味構造を合成してつくられるコマンドである。

コマンド合成においては次の規則が適用される。つまり、MAKE コマンドでは、更新用属性として、ZAI または SEI を選び、そのほかを条件とする。SEL コマンドでは、疑問の対象となる属性つまり、「いくつ」、「なに」などの疑問詞に係る属性が検索の対象となり、ほかの属性は条件部に記述する。TEST コマンドでは、すべての属性が条件となる。

コマンド生成の例を図8に示す。

一行目に文節に分解された文を示す。二行目は、各文節の意味品詞と意味単語を表わしている。三行目以降に、係り受けにより決定される属性名と属性値、および最終的に生成されるコマンドを示す。

a の説明をする。「アスファルトタイルは」に関しては、意味品詞、意味単語がそのまま、属性名 ZAI、属性値 ASPHALT-TILE となる。また、「摩耗に」と「耐える」の係り受けにより、属性名は SEI (性能)、属性値は GA 4 (耐摩耗性) となる。同じく、「相当」と「耐える」の係り受けにより属性名は TOU (等級)、

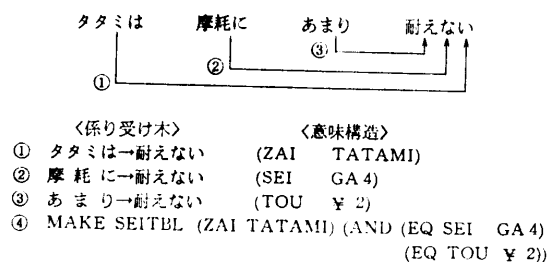


図7 コマンドの生成過程

Fig. 7 Command generate process.

属性値は ¥ 8 (8 ぐらい) となる。平叙文なので、コマンド名は MAKE となる。SEI, ZAI, TOU の3つの属性が使われているので関係名は SEITBL となる。3つの属性のうち2つを条件とし、1つを更新用属性として、最終的なコマンドを生成する。

a の文は、b と c のように言い換えることも可能である。文が変わっても、意味が同じ場合は、同じコマンドが生成される。

d は、WHAT 型疑問文なので、SEL コマンドが生成される。「材料は」「ありますか」に関しては、意味単語はない。「なにが」と「ありますか」の係り受けにより、WHAT 型疑問文であると判定される。属性は、BE と ZAI なので、関係名は BETBL となる。このコマンドを実行すると、関係名 BETBL の属性名 BE の属性値が YUKA である組をさがし、該当する組の ZAI を返す。

e も、「いくつ」と「ですか」の係り受けによって WHAT 型疑問文と判断され、SEL コマンドが作成される。ZAI, SEI, TOU の3つの属性を使っているので、関係名は SEITBL となる。

f は、YES/NO 型疑問文なので TEST コマンドが生成される。ZAI, SEI, TOU の3つの属性は、すべて条件部におかれる。このコマンドは属性名 SEI の属性値が GA 4 で属性名 ZAI の属性値が TATAMI でかつ、属性名 TOU の属性値が5以上である組をさがし、該当する組があれば T を返し、なければ F を返す。

g も、YES/NO 型疑問文なので、TEST コマンドが生成される。

コマンドの実行管理を行うのが検索更新である。すなわちデータベースの検索、更新、追加などを行い、結果を出力する。

6. システムの実動化

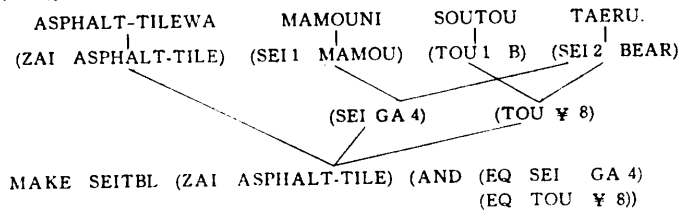
本システムは日本語を使った、建築材料に関する質問応答システムであり、現在は、材料の性能および材料の使用部位 (以下 BE と記す) に関する質問応答が可能である。

利用者は日本語で平叙文と疑問文を入力して、データベースを操作することができる。

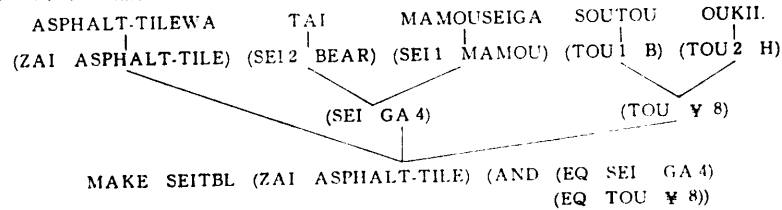
平叙文を入力することにより、データベースの更新追加を行える。正常に操作が完了すると「OK」と応答する。

疑問文を入力すれば、データベースの検索ができ

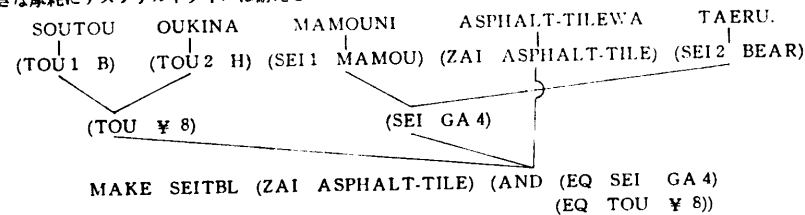
a. アスファルトタイルは摩耗に相当耐える。



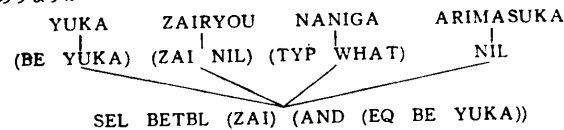
b. アスファルトタイルは耐摩耗性が相当大きい。



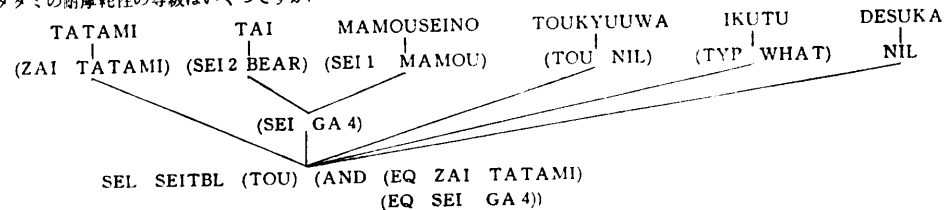
c. 相当大きな摩耗にアスファルトタイルは耐える。



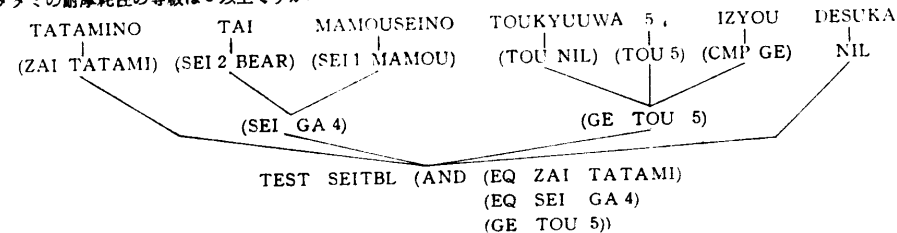
d. 床材料にはなにがありますか。



e. タタミの耐摩耗性の等級はいくつですか。



f. タタミの耐摩耗性の等級は5以上ですか。



g. タタミは床材料ですか。

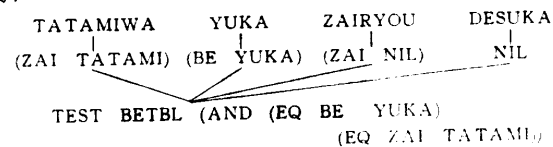


図 8 コマンド生成例

Fig. 8 Examples of Command Generaion.

る。YES/NO 疑問文の場合は、質問の内容が正しければ「YES」と応答し、正しくなければ「NO」と応答する WHAT 型疑問文では、質問の内容を答える。

日本語はローマ字でタイプするが、特に分かち書きされていないなくても、自動的に文節単位への切り分けを行う。また意味不明の部分があるときは、意味不明の箇所を提示し、再入力を要求する。

次にいくつかの実例をあげる。

1) ガムシートは摩耗に相当耐えるか。

@GUM-SHEETWA MAMOUNI SOUTOU
TAERU.

OK

材料性能に関する平叙文である。もし、ガムシートの性能に関する情報がデータベースにあれば、データベースを更新し、なければ、追加する。

2) ガムシートは摩耗に相当耐えるか

@GUM-SHEETWA MAMOUNI SOUTOU
TAERUKA.

YES

材料性能に関する YES/NO 型疑問文である。データベースを検索した結果、質問内容が正しかったので、応答は「YES」となる。

3) 大理石は屋根材料か

@MARBLEWA YANE ZAIRYOUKA.

NO

材料の BE に関する YES/NO 型疑問文である。大理石は主に床材料なので、応答は「NO」となる。

4) ガムシートの耐熱性の等級はいくつだ。

@GUM-SHEETNO TAINETUSEINO
TOUKYUUWA IKUTUDA.

4

材料性能に関する WHAT 型疑問文である。性能は 0~9 にグレード分けしてあり、9 に近いほど、性能が良いことを示す。データベースにはガムシートの耐熱性の等級が 4 で登録されているので、応答は「4」となる。

5) 床材料はなんだ。

@YUKAZAIRYOUWA NANDA.

MARBLE HARD-BOARD COLK-
TILE WILTON-CARPET RUG
TATAMI ASPHALT-BLOCK
TERRAZO-BLOCK CLINKAR-
TILE MOSAIC-TILE GUM-TILE

LINOLUEM-TILE LINOLUEM-SHEET
GUM-SHEET

材料の BE に関する WHAT 型疑問文である。床材料としてデータベースに登録されている材料がすべて出力される。

6) アスファルトブロックはどこで使用されるか

@ASPHALT-BLOCKWA DOKODE SIYOU-
SARERUKA.

YUKA

これも、材料の BE に関する WHAT 型疑問文である。5) が材料名を質問しているのに対して、6) は BE を質問している。

7) タタミは水に耐えない。

@TATAMIWAMIZUNITAENAI.

OK

分かち書きされていない文の例である。システムは最長一致で文節の切り分けを行う。

8) 大理石はどこで使用されるか

@MARBLEWA DOKODE SITOU SARERUKA.

IMI HUMEI SITOU SITOU

意味不明の例である。システムは意味不明の単語を示して、再び、文を入力することを要求する。

図 9 に、さらにいくつかの実行例を示す。

7. あとがき

以上のような、日本語による、建築材料に関する質問応答システムの研究開発をした。

@LINOLUEM-SHEETWA DOKODE SIYOU SARERUKA.

YUKA

@MOSAIC-TILENO TAIMAMOUSEINO TOUKYUUWA OUKIKA.

NO

@MOSAIC-TILENO TAIMAMOUSEINO TOUKYUUWA IKUTUDA.

8

@MOSAIC-TILENO TAIMAMOUSEINO TOUKYUUWA SOUTOU TIISAI.

OK

@MOSAIC-TILENO TAIMAMOUSEINO TOUKYUUWA IKUTUDA.

¥2

@LINOLUEM-TILENO TAISUISEINO TOUKYUUWA IKUTUDA.

6

@LINOLUEM-TILEWA DOKODE SIYOUSARERUKA.

YUKA

@LINOLUEM-TILENO TAISUISEINO TOUKYUUWA TIISAI.

OK

@LINOLUEM-TILENO TAISUISEINO TOUKYUUWA IKUTUDA.

¥L

@ASPHALT-TILEWA MIZUO TOUSUKA.

YES

@GUM-SHEETWA YANEZAIROYUDA.

OK

@YANEZAIROYUWA NANDA.

GUM-SHEET

図 9 実行例

Fig. 9 Example of conversation.

ほかの形式的なコマンドに比べ、より非手続き的な表現が可能である。また、日常言語に近いため、覚えやすく分かりやすい。

自然言語インタフェースの場合、指定に柔軟性があることが長所になるが、無制限に柔軟ではないので、利用者はどの程度の柔軟性をもつかの判断に迷うという欠点にもなる。そのため、会話を援助する機能（ミス入力の指定、ミス入力の訂正等）の充実が必要であり、今後の課題と考える。

現在は、センタシステム上の制限で、ローマ字文を入力しているが、カナ文字あるいは漢字で入力できればより使いやすくなると思える。

単語数は、詞に 160 語、辞に 140 語程度もつ。材料数は 20 程度、性能の種類としては 20 ほどある。計算機および使用言語は、東京大学大型計算機センター HITAC 8800/8700 と HLISP である。

現在は実験システムの段階であるが、データベースを充実させることなどにより実用も可能になると思える。

謝辞

最後に、貴重な助言とご指導をいただいた東京理科大学の三重野博司助教授と小池慎一講師に深く謝意を表します。

参 考 文 献

- 1) Woods, W. A.: Progress in Natural Language Understanding-An Application to Lunar Geology, Natnl. Comp. Conf., 28 pp. 441-450 (1973).
- 2) Plath, W. J.: The REQUEST System, IBM Res. Tech. Report, RC 5604, p. 20 (Aug. 1975).
- 3) Mylopoulos, J. et al. TORUS: A Step towards Bridging the Gap between Data Bases and the Casual User, Inf. Syst., Vol. 2, pp. 49-64 (1976).
- 4) 藤崎哲之助他: データベース照会システム「ヤチマタ」と名詞句データ模型, 情報処理学会論文誌, Vol. 20, No. 1, pp. 77-84 (Jan. 1979).
- 5) Astrahan, M. M.: SYSTEM. R, ACM Transaction on Data Base Systems Vol. 1, No. 2, pp. 97-137 (June 1976).
- 6) Codd, E. F.: Seven Steps to Rendezvous with the Casual User, pp. 179-200, North-Holland (1974).
- 7) Hendrix, G. G. et al.: Developing a Natural Language Interface to Complex Data, ACM, Trans. on Data Base System, Vol. 3, No. 2 (1978).

(昭和 55 年 7 月 30 日受付)

(昭和 56 年 4 月 27 日採録)