HOKKAIDO UNIVERSITY

Title	Fuzzy質問応答システムの構成及び機能について
Author(s)	橋谷, 敏郎; 佐藤, 義治; 遠藤, 経一 他
Citation	北海道大學工學部研究報告, 85, 115-121
Issue Date	1977-09-16
Doc URL	https://hdl.handle.net/2115/41436
Type	departmental bulletin paper
File Information	85_115-122.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北海道大学工学部研究報告
第85号（昭和52年）

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 85 (1977)

Fuzzy 質問応答システムの構成及び機能について

橋谷敏郎* 佐藤義治*
遠藤経一* 河口至商*

（昭和52年3月31日受理）

A Construction and Functions in the Fuzzy Question-Answering System

Toshiro HASHIYA Yoshiharu SATO
Tsunekazu ENDO Michiaki KAWAGUCHI

(Received March 31, 1977)

Abstract

The ordinary question-answering systems are regarded to have the following inference mechanism in which the assertions are either true or false and the action-rules are either valid or invalid.

This mechanism is not available for cases including inexact informations and for practical applications.

This report discusses a fuzzy question-answering system with a medical data base. This system is constructed from a natural language analyser, flexible pattern-matching routine, and searching routine for the manipulation of the fuzzy data base. The language analyser is based on the case grammar by C. J. Fillmore and this grammar may be suitable for the Japanese. And we use two inference rules based on Modus Ponens and the equivalent and analogical inference algorithms.

Then, this system can manipulate non-precise and fuzzy information and the mechanism of this system will be usefull to approach to human question-answering behavior.

1. まえがき

質問応答システムは人工知能の分野において，主要なテーマの一つと考えられている。質問応答システムとは，与えられた質問に対して，それを解析（理解）し，さらに必要な推論を行い，解答を作成するという，計算機による知的情報処理システムの一つである。このようなシステムを構成する際に問題となる点は，質問をどのような言語形式で与えるかということ，さらに解答を作成するときに用いる推論機構およびそれに附随するデータベースの形態等である。言語形式については自然言語形式であることが最も望ましいものであり，現在までに作られているシステムはほとんど自然言語を対象としている。ところが，自然言語を対象とした場合，その意味構造

* 工学部情報数理工学第一講座

には，あるあいまいさが含まれて来る。また推論形式においても，人間の日常生活における思考過程（推論）には多分にあいまいさを含んでいることがある。従来作成されているシステムでは，このようなあいまいさを除去した体系の中で構成されているものが多い。しかしながら，より現実的な問題に対応できる質問応答システムを構成しようとするとき，ある意味のあいまいさをもった情報をも処理できるシステムであることが必要とされる場合がある。この意味において，ここでは，あいまいさを考慮に入れたFuzzy質問応答システムの構成と機能について報告するものである。

一般に質問応答システムが所有すべき能力として，現在つぎのようなものが要求されている。

(i) 自然言語によるシステムとの対話が可能であること。

(ii) 複雑な情報の効率的蓄積および運用ができること。

(iii) 種々の推論および分析手段を保有していること。

実際に，与えられた質問に答えるためには，それに対する知識を所有していなければならない。従ってそれらを効率的な内部形式で保有し，かつ効率よく運用することが必要である。この種の問題は，従来情報検索の分野およびデータベースとして研究されているが，質問応答システムの場合においても，効率的運用の可能なデータベースを保有していなければならない。また質問に対して直接解答ができる情報が存在しないときでも，データベースの知識を用いて推論を行ないそれに対する解答を作成する能力は特に重要であり，豊富な推論手段をもっていることが必要である。本報告における Fuzzy 質問応答システムは，以上の点を考慮して，自然言語解析のための文法，あいまいさを含む情報をも蓄積しているデータベース，およびそれを用いての推論を行ないくつかの手続きにより構成されている。またここではデータベースとして主に医学情報をとり上げている。

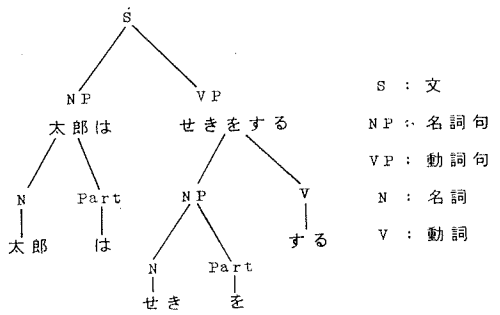
2. 格文法による自然言語解析

システムとの対話が自然言語によってできることは，その使用者にとって最も便利であろう。定められた形式，例えば論理式や関係等での対話は，それら種々の規則を使用者は知っていなければならず，そのシステムは普通の使用者には使い難いものであろう。電子計算機での自然言語理解という研究分野は近年非常に発展して来ており，特に T. Winograd の SRDHLU の成功以来新しい局面を迎えている。これらはもちろん人間の言語解析能力を実現しているわけではないが，今後種々の分野に影響を与えるものと思われる。

本研究の Fuzzy 質問応答システムにおいては，言語学者 C. J. Fillmore により提案された格文法に基づき日本語文解析のための言語解析ルーチンを構成した。この分野において，自然言語解析のために使われる文法理論としては，N. Chomsky の変形文法を基礎としたものが多い。例えば

「太郎はせきをする。」

という文は，主部である「太郎は」と，述部である「せきをする」から成り，更に主部は「太郎」（主語）と「は」（付属語）に分解され，述部は「せきを」（名詞句）と「する」（動詞）に分解され，述部中の名詞は更に「せき」（名詞）と「を」（付属語）に分解される。これを木構造と



図—1 例文の木構造表現

して図示すれば 図-1 の様になる。

我々は次の様な理由により変形文法に基づく解析は行なわなかった。

(1) 変形文法は構文解析には適しているが，意味解析には適していない。

すなわち，自然言語文の表現している意味内容の理解のために，複雑な処理を要求する変形文法は，言語解析手続きを繁雑かつ非能率なものにしてしまう。

(2) 複雑な文の木構造においては，枝の数が莫大なものになってしまう。

自然言語文の解析結果としての木構造はリスト形式で計算機の内部に格納されるが，枝の数が多くなればその処理は困難になってしまい，できるだけ枝の数が多くならない事が望ましい。

(3) 変形文法は日本語文の解析のために特に適した文法とは思えない。

日本語は比較的名詞句が互いに独立しており，英語の様に解析のための文法として変形文法を用いる事は有利ではない。格文法はその点日本語に適した文法である部分が多い。

格文法は自然言語文について，その文中の動詞が必要とする格を必須格とし，動詞に付与する事は可能であり文中に存在する事は許されるがなくても良い格を自由格として，それらから成る格構造により自然言語文を解析する文法理論である[1]。本システムでは取扱う情報の特徴及び複雑な文の処理は現段階では行なわなかったので許される。格は次の四種類のみとした。

必須格……主格 (SUB)，目的格 (OBJ)

自由格……場所格 (LOC)，時間格 (TIME)

言語解析ルーチンは，日本語での入力文を受け取ると，最初に入力文の単語の辞書引きを行ない入力文中の全ての単語の品詞分類を行なう。次に「の」を除いた他の格助詞の部分で入力文を区切り，区切られた部分を名詞句処理ルーチンに渡して，実際に正しい形容詞の修飾がなされているか，正しい語順かどうか等の名詞句処理を行なう。最後に処理されたそれぞれの名詞句はそれに付いていた格助詞の種類により文中の動詞に対する格構造の決定を行ない，それらが全て成功すれば入力文の処理が終り，格構造に基づく解析結果をシステムは獲得する。本システムで使用を許している格助詞の種類は 表-1 の通りである。

表—1 格助詞の種類

は	SUB
が	SUB, OBJ
を	OBJ
から	LOC, TIME
より	TIME, LOC
に	TIME, LOC
の	
へ	LOC

これにより，例えば

「かわいい子供が昨夜より赤い発疹を口の中に出した」

という文は，次の様に分析される。

格助詞「が」までの部分が最初に名詞句処理されるが「子供」は上位概念として「人間」を持つ物質名詞で，属性「特徴」として「かわいい」という特徴を説明する形容詞による修飾が可能である。格助詞「が」はこの名詞句が主格になる事を指示して解析結果は内部形式として

(SUB (TOK KODOMO) (TOK KAWAII))

の形式になる。次に「昨夜より」は「昨夜」という機能名詞は属性として時間を持ち，「より」は時間格に対応する格助詞だから分析結果は

(TIME (TOK SAKUYA))

となる。動詞「出した」は過去形であるがシステムの動詞辞書では語尾変化については現在基本的な時制か疑問か肯定かの区別に関する情報のみを所有している。

以上より，上の例文は次図のような分析結果となる。これは変形文法より，木構造として見た場合の枝の数も相当少なくてすむという事が想像されよう。

例文の分析結果のリスト表現

(VERB (TOK DASU) (TENSE KAKO))
(SUB (TOK KODOMO) (MOD KAWAII))
(OBT (TOK HASSHIN) (MOB AKAI))
(TIME (TOK SAKUYA)) (LOC (TOK NAKA) (MODIN KUCHI))
(DEG SURELY))

以上の様に入力文の解析は行なわれるが，現在本システムでは複雑な入力文の処理，例えば接続詞でつながった複合文，関係代名詞，連体修飾，省略語等の処理は行なっていない。

これらに関する能力の向上は次の段階の仕事であるが，以上の制限はあるものの，この程度の自然言語解析能力でも質問応答システムとして，ある程度満足の行く対話能力であると思える。

3. データベースの機構と推論手法

本節では最初に，取扱う情報の形式化及びデータベースの機構について説明する。

前節で説明した様に，本システムでは日本語文による入力を格文法に基づき主解析し，内部形式に変換するのであるから，データベース内に保有する情報の表現形式も同様の形式に依るのが有利である。我々はあいまいな情報の取扱いのために，L. A. Zadeh による Fuzzy 論理を用いたのであるが，Fnzzy 論理は次の様なものである。

ある言明 E に対してその真理値は又は確定度とも解釈される fuzzy grade と呼ばれる区間の間 [0, 1] の実数が，E に対して一通りに付与され，それは次の規則に従う。

E_1, E_2 の fuzzy grade をそれぞれ $S(E_1)$, $S(E_2)$ とすれば

(i) E_1 の否定 ($\neg E_1$ で表わす) は

$$S(\neg E_1) = 1 - S(E_1)$$

(ii) E_1, E_2 の積 (and), 和 (or) は

$$S(E_1 \,\&\, E_2) = \min(S(E_1), S(E_2))$$
$$S(E_1 \vee E_2) = \max(S(E_1), S(E_2))$$

(iii) 条件的含意　$E_1 \supset E_2$ ($\supset$ はならばである) は

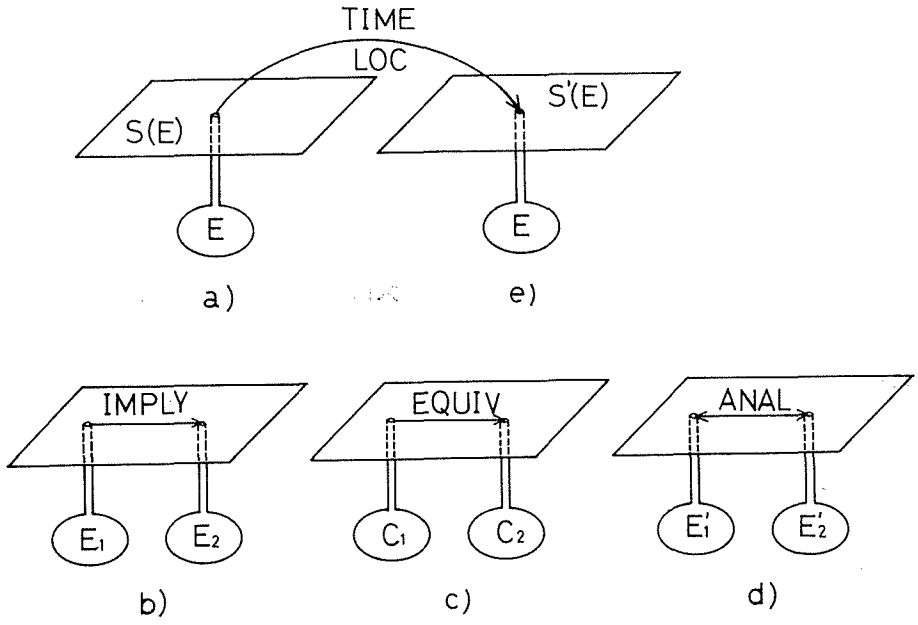
$$S(E_1 \supset E_2) = \max(S(E_1), 1 - S(E_2))$$

として，それぞれ定義される。我々は fuzzy grade が確定した言明を fuzzy 事実，又条件的含意を fuzzy 知識と呼ぶ事にする。

データベースは以上の様な fuzzy 事実と fuzzy 知識の集合及びそれらに対する補助的知識からなるセマンテック・ネットワークである。

図-2 により本システムのデートベースとしてのセマンテック・ネットワークについて，その特徴を説明する。

図-2 において，a) は fuzz 事実として事象 E が格納されている事を表現している。b) は S 関係 IMPLY により前提 E_1 と結論 E_2 から成る条件的含意，すなわち fuzzy 知識である。又 c) は S 関係 EQUIV による同値概念の知識，d) は S 関係 ANAL による類似概念の知識であるが，これらは推論手法の説明の時に説明する事にする。e) は次に説明するが fuzzy 事実 E′ が場所変位，時間的推移によりその fuzzy grade が変化する事を表現している。すなわち fuzzy grade は事象の起こった場所，起こった時間等に依存する相対的真理値であるものと我々は解釈した。



図—2 S 関係によるセマンテック・ネットワーク

さて，我々は取扱う情報として医学情報特に小供の病気を選んだが，一般に病気は次の様な要素から特徴づけられるであろう。

(i) 病気の引起こす種々の症状及びその出現の程度

(ii) 症状の出現する時間的特徴（出現時間，継続期間）

(iii) 症状の出現する場所的特徴（出現部位）

上の 3 つにより病気を特徴づける事ができるが，(ii), (iii) については，格としてそれぞれ時間格，場所格に対応するものであり，出現の程度等より fuzzy grade が決定され，医学情報の fuzzy 事実としての形式化がなされる。例えば，子供が麻疹（はしか）にかかっているとすれば，

「子供は発疹期に高い熱を出す。」

という言明は

(HEIJOU
((UERB (TOK DASU)) (SUB (TOK KODOMO)) (OBJ (TOK NETSU) (MOD TAKAI))
(TIME (TOK HASSHINKI)) (DEG LIKELY)))

の様に，格構造に基づいた内部形式でデータベース内に格納されている。その他の情報も 図-2 からわかるようにそれぞれ S 関係に基づいたマーカーが付与されてデータベースに格納されている。

次に本システムで用いられる推論について説明する。機械的推論アルゴリズムとして従来まで提案されたものの内で最も有名なのは，J. A. Robinson による導出原理である。

しかし実際の人間の推論はあいまいさ，不確実さが伴なうので，厳密な論理的推論能力のみより，発見的な演繹を追求すべきである。それ故情報を前節の文法に基づく内部形式でデータベースとして保持して，効率的パターン・マッチングを主体とする推論手法を本システムでは採用した。基本的演繹能力としては次に述べる一般化された三段論法 (modus ponens) がある。

$$\frac{A \quad A \supset B}{B}$$

のような図式において，Aはfuzzy事実，A⊃Bはfuzzy知識である。質問Bに対する情報がデータベース内に存在してはいないが，AとA⊃Bが存在しているならば，一般に三段論法は

「Aが成立ち，A⊃Bが成立てば，Bが成立つ。」

という推論規則であるから，上の図式により質問Bへ解答する事が可能である。fuzzy事実又は，未知の情報に関する問合せの質問を除いて，上の三段論法は本システムの基本的推論能力として使用される。現在推論の深さは2段階までであるが，大部分の質問に関して十分である様に思われる。

又，三段論法の他に，次の様な推論手法を採用する。fuzzy事実とfuzzy知識の他にデータベースはS関係 EQUIV, ANAL 等で定義される知識も所有している。例えば子供の病気である麻疹について，発病期間は前駆期，発疹期，回復期の順に区切られるが，「前駆期の後」という名詞と「発疹期」という名詞は同値であり，又前駆期における症状は風邪の症状に類似している。推論がうまく行かなかった場合，推論ルーチンはこれら同値概念による質問文の変形，類似的情報の探索等を行なう。特に後者はJ. R. Carbonellによる Functional Inference の一種であり，本システムの問題に密接な関係を持っており，今後この推論手法の発展が望まれる。質問に対して生成される解答に付与されるfuzzy gradeは前に述べた規則に従がうが，三段論法が重復して用いられたり，類推並びに時間，場所的知識を使用した時は，それに対応して変更される。

4. あとがき

以上あいまいさを伴なう情報を取扱う自然言語解析能力と演繹的推論能力を有するFuzzy質問応答システムの機構について述べた。結論として次の様な点が言えよう。

(1) 自然言語解析，特に日本語解析には格文法が非常に有効である。
だが取扱う単語数の制限，深い係り受け，関係代名詞，連体形等の処理を本システムでは行なわないので複雑な入力は処理できない。より一層の解析能力を増すための努力が必要である。

(2) Fuzzy情報としてデータを形式化する事は本システムの最も重要な課題であったが，実際の従来まで取扱えなかった領域まで取扱える幅広さを得た。

(3) 推論能力としては，三段論法の他に，本システムは同値，類似等の知識を用いての推論能力を提案したが，更に効果的な改良が必要である。

(4) Fuzzy情報として本システムは医学情報を対象としたが，今後この分野は人工知能的観点からの研究がなされるべきである。

以上の様に基本的には初期の目的は達成されたが，今後更に上述の課題のための検討を現在行なっている。

参 考 文 献

1) Fillmore, C. J. (1968) The case for case, Univesals in Lingustic Theory, Bach and Hanus, eds.,
2) 橋谷敏郎, 沢村 一, 前田 隆, 河口至商 (1976) Fuzzy 情報を取扱う質問応答システムについて, 昭和51年度情報処理学会第17回全国大会.
3) 橋谷敏郎 (1977) Fuzzy 質問応答システムの一つの試み, 昭和51年度北海道大学大学院工学研究科情報工学専攻修士論文.
4) Kunii, T. (1976) DATAPLAN: An Interface Generator for Data Base Semantics, Information Science, Vol. 10, 279-298.
5) Simmons, R. F. (1973) Sematic Networks: Their computation and use for understanding Sentences, in Computer Models of Thought and Languuges, Schank (R. C. & Colby, K. M. eds.).
6) Winograd, T. (1972) Understanding Natural Language, Academic Press
7) Zadeh, L. A. (1975) Fuzzy logic and approximate reasoning, Synthese, Vol. 30, 407-428.