

# 言葉で考えるコンピュータの実現にむけて

## An Approach to Computer Thinking with Words

石川 勉<sup>\*1</sup>

Tsutomu Ishikawa

<sup>\*1</sup> 拓殖大学工学部情報工学科

Department of Computer Science, Faculty of Engineering, University of Takushoku #1

This paper introduces an approach to realize “Computer Thinking with Words,” which can comprehend and solve the problems represented with natural language, and then answer them using the language. This computer, aiming at some applications in our daily life such as dialogue systems, consulting systems and robot’s brain in the future, makes use of common knowledge in addition to problem-specific knowledge for solving given problems. They are represented using a new scheme based on predicate logic, in which Japanese ordinary words are used as knowledge elements basically. In this article, we present an inference mechanism which can operate based on the scheme and obtain approximate solutions even when problem-specific knowledge is incomplete, a translation method from Japanese sentence into the scheme and acquisition of common knowledge concerning to words from machine-readable dictionaries.

### 1. はじめに

我々は、人間の判断に近い柔らかな処理をコンピュータ上に実現することを目指し研究を進めてきた。具体的には、この実現にむけ当面は、「常識を活用すること」および「概略解を狙うこと」に着目し、まず「常識」を構成するための大規模知識ベース（以下、常識知識ベースと呼ぶ）および「常識」を活用する推論技術について研究してきた。これまで、前者については、常識知識ベースの基本部分となる単語に関する知識ベース（以下、概念ベースと呼ぶ）の構築を進め、言葉の意味の類似性判定を主目的とした 26 万語規模の概念ベースを構築してきている [Nguyen 02]。また、後者については、欠落した知識を類似の知識で補完して概略的な解を得る推論法（以下、概略推論法）を提案し、実験システムによりその基本動作を確認している [Nguyen01]。さらにこれらの有効性を確認するため、具体的な応用である自由対話システムの開発を進め、簡単な質問応答（問題解決）機能を持つ実験システムの製作を行なった [小林 01]。

我々はこれらにより冒頭の目標に対して一定の足がかりは得られたと判断し、さらにこれらを統合、発展させた「言葉を理解し、言葉で考え、言葉で答える」コンピュータ（以下、「言葉で考えるコンピュータ」と呼ぶ）の実現という、より具体的な目標を設定した。ここでは、まずこのコンピュータの構想について述べ、次に、その核となる、言葉ベースの知識表現法、それに基づく推論法、自然言語文の上記表現への自動変換法、この変換法を利用した国語辞書からの常識知識の獲得法に関する、現在までの検討状況について述べる。

### 2. 「言葉で考えるコンピュータ」の構想

人間あるいは文書等、外界とのインターフェイスから内部的な処理まですべて言葉をベースとするコンピュータを意識している。言葉もっと広く捉えれば、言語は古来から多くの哲学者がその本質、意味、思考との関係等について議論し続けている厄介な代物である [野本 02]。「これを機械に分かるように厳密に表現し、それをもとに考えさせ、何らかの結論を出させる」などというこ

とは夢のまた夢である。従って、ここでは現在のコンピュータ化可能な A 技術、自然言語技術を前提に、冒頭でも述べたように厳密性（推論で言えば健全性）、正確性は捨て、何らかのそれらしい解を得るメカニズムを追求する。勿論、人間（外界）とのインターフェイスは自然言語が前提である。らしい解であるからそれを頭から信用して動くものへの応用は不可能であり、その解を人間が評価できるような応用（人間の介在が不可欠）が前提となる。例えば、対話機能を有する 癒し系ロボットの脳、さらには広くヒューマンインターフェイスの柔軟化のための諸ツールが考えられよう。

らしい解であれ何であれ、解を出す（考える）には推論メカニズムが必用であり、そのための知識表現形式を定める必要がある。これには、フレーム、意味ネット、ルールベース、述語論理あるいはそれらの変形、最近ではオントロジ記述への適用性から DL [兼岩 03] 等も注目されているが、ここでは、推論処理だけでなく自然言語処理との適合性、実現容易性等々を考慮し、基本的には 1 階述語論理をベースとする知識表現法をとる。ただし、これにより表現できる自然言語文の範囲はごく僅かなため、それを扱う推論メカニズムの実現可能性を考慮しつつ、その拡張を図る。

また、「考える」には当然その基となる知識が必要になるが、基本的には従来のエキスパートシステムと同様にその問題対応の知識（問題解決規則等）はユーザ側で用意しておく。しかし、日常的な問題を想定してこの知識を書ききることは不可能に近い。システム側でそれを補完する常識知識を収容しておく。この常識知識は電子化文書から自動獲得することを想定する。従って、これら知識の完全性は期待できず（TM S等の知識の一貫性管理システムの利用も考えられるが、それだけで解決できる不完全性のレベルでないことは当然予想される）、何らかの知識精練さらには推論における知識のフォールトトレランスも不可欠になる。また、このような知識を前提とした推論であるから、従来の厳密な推論メカニズムの適用は無意味であり、これまで提案している概略推論を発展させ適用する。さらにこの推論では、可能性のある多くの解が導かれることが想定されるためその評価が必用となることは勿論のこと、推論に行き詰まったときにその方向を示すメタ知識（価値観的）も必用となる。図 1 にこのコンピュータの構成イメージを示す。

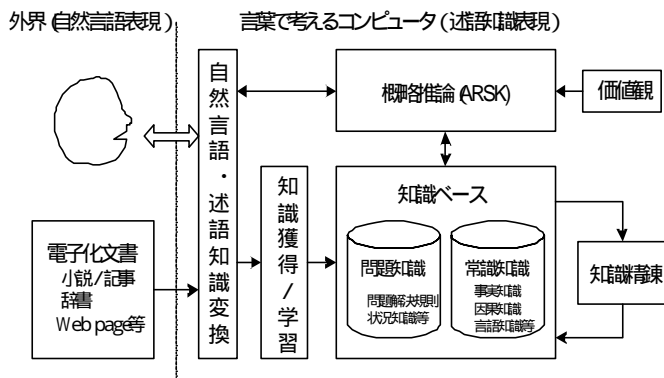


図1 言葉で考えるコンピュータの構成イメージ

### 3. 言葉ベースの知識表現法

1階述語論理をベースとするが、さらに以下の条件を考慮した知識表現形式とする。

- ・対話文等の比較的単純な文 (単文および一部の複文) をその表現対象とする。
- ・電子化辞書等の利用可能なツールを用いて、自然言語文から容易に変換できる。
- ・その形式で表現された知識の意味が一意に定まる。
- ・日常的問題を意識し、知識の構成要素は単語をベースとする。

#### 3.1 基本的な表現法

1つの文 (命題) をどう表すかについては、文全体を1つの複雑な素式 (新形式) で表現する方法と複数の単純な素式 (一般的な1階述語論理形式) の連言で表現する方法 (素式間は共通の定数で意味的に結合) が考えられる。ここでは、推論処理の複雑化を避けるため後者の方法をとる。すなわち、文をできるだけ単純な素式の連言で表すようにする。

以上を考慮し、素式およびそれを構成する述語、引数 (項) は以下のように設定する。

素式: 述語(ラベル:引数, ラベル:引数, ..., ラベル:引数)

述語: 動詞

・名詞, 複合語 (名詞・名詞... 名詞), 名詞句 (名詞 + 助詞 + ... 助詞 + 名詞)。ただし、最後の名詞以外は定数, 変数でも可)

・形容詞, 形容動詞 (これらを修飾する副詞がつく場合を含む)

引数: 名詞, 複合語, 名詞句 (上記と同じ), 変数, 定数  
すなわち, 知識構成要素は基本的に単語とする。

#### 3.2 具体的な表現法

素式の述語は、それにより表現しようとする文の述部の品詞により以下のように決定する。

述部が動詞のとき 述語はその動詞で引数はその深層格。  
述部が名詞のとき: 「~は~である」の文型の場合であり、述語は名詞。

述部が形容詞のとき 述語は形容詞。

以下、これらの場合毎について示す。

##### (1) 述語が動詞の場合

表1. 主な深層格とラベル

| 格  | 概念関係子     | ラベル | 概要        |
|----|-----------|-----|-----------|
| 主体 | Agent     | agt | 動作を行う主体   |
| 対象 | Object    | obj | 動作, 思考の対象 |
| 目標 | Goal      | gl  | 動作の終点     |
| 源泉 | Source    | su  | 動作の始点     |
| 場所 | Place     | plc | 事象の成立する場所 |
| 場面 | Scene     | sce | 事象の成立する場面 |
| 道具 | Implement | inp | 動作の手段, 道具 |
| 材料 | Material  | mat | 材料        |

表2. 動詞の態様

| 態様 | 記号 | 表現      |
|----|----|---------|
| 否定 | !  | ~しない    |
| 過去 | *  | ~した     |
| 希望 | w  | ~したい    |
| 可能 | c  | ~できる    |
| 提案 | s  | ~すべき    |
| 許可 | m  | ~してもよい  |
| 推測 | p  | ~かもしれない |

引数のラベルは深層格毎に定めることとし、これにはE D R電子化辞書における概念関係子を用いる。表1に主な深層格とラベルを示す。さらに、時制, 否定, 様相 (動詞の態様) についても表2のような単純なものについては表現可能とする。以下に、基本的な表現形式と表現例を示す。

[表現形式] 動詞の態様を表す記号 動詞の終止形 (ラベル: 深層格, ..., ラベル: 深層格)

[例] 「太郎は朝食を食べる」 食べる(agt:太郎, obj:朝食)  
「太郎は本を読まない」 !読む(agt:太郎, obj:本)

##### (2) 述語が名詞の場合

「AはBである」という文型を想定する。ここで、A, Bは名詞である。この場合、一般にBが普通名詞の場合にはBがAの性質を、固有名詞の場合にはAがBの性質を表す。従って、基本的な表現形式を、ラベル inst を用い以下のように設定する。

[表現形式] Bが普通名詞の場合: B(inst:A)

Bが固有名詞の場合: A(inst:B)

[例] 「太郎は医者である」 医者(inst:太郎)  
「犯人は太郎である」 犯人(inst:太郎)

なお、「AはBである」という文型におけるAとBの関係は多様であり、上記以外にも多くの関係がある。これらの詳細については別途報告する予定である。

##### (3) 述語が形容詞の場合

「AはB」、「A<sub>1</sub>はA<sub>2</sub>よりB」、「A<sub>1</sub>はA<sub>2</sub>がB」の文型を対象 (Bが形容詞) とし、ラベル sbj, r-sbj, p-sbj を用いてそれぞれ以下のように表現する。

[表現形式] 「AはB」 : B(sbj:A)

「A<sub>1</sub>はA<sub>2</sub>よりB」 : B(sbj:A<sub>1</sub>, r-sbj:A<sub>2</sub>)

「A<sub>1</sub>はA<sub>2</sub>がB」 : B(sbj:A<sub>1</sub>, p-sbj:A<sub>2</sub>)

[例] 「花子は美しい」 美しい(sbj:花子)  
「花子は良子より美しい」 美しい(sbj:花子, r-sbj:良子)  
「花子は目が美しい」 美しい(sbj:花子, p-sbj:目)

なお、述語または引数が名詞節となる場合あるいは形容詞や動詞(連体形)を含む名詞節となる場合については、前節の考え方にもとづき素式の連言で表現する。具体的には、名詞節ある

いは名詞句で表す内容を定数  $S$  で表し、元の文を表す素式と  $S$  を用いて名詞節等を表現する素式群の連言で表す。

【例】 花子は太郎が経営している会社の社員である

$S$  の社員(inst:花子) 会社(inst:S)  
経営する(agt:太郎, obj:S)

【例】 太郎は汚い机の上のきれいな花を見ている

見る(agt:太郎, obj:S<sub>1</sub>) S<sub>2</sub>の上の花(inst:S<sub>1</sub>)  
きれい(sbj:S<sub>1</sub>) 汚い(sbj:S<sub>2</sub>) 机(inst:S<sub>2</sub>)

#### 4. 知識の不完全性を許容する推論法

ベースとする概略推論法は、以下のように定義される。

$K \vdash G$   
 $K \cup K_s \vdash G$   
 $K_s = \{k_i \mid \exists k \in K, k \approx_{\nu} k_i\}$

すなわち、もとの問題知識  $K$  からはゴール  $G$  が証明されない場合に、 $K$  中の知識  $k$  に類似する知識  $k_i$  の集合  $K_s$  を加え  $G$  を証明しようとするものである。類似知識  $k_i$  の検出はそのときの推論状況  $V_i$  を考慮して動的に行い、知識間の類似性は知識を構成する単語の意味的な類似性をもとに判定する。この類似性判定には概念ベースを用いる。

推論の方式としては、類似知識の適用の容易性を考慮して、基本的には後ろ向き推論の SLD (Selective Linear resolution for Definite clauses) 法を用いている。SLD 法では、知識ベースと証明したいゴールの否定から導出手続きを用い空節を導く。空節が導かれれば証明が成功したことになる。本推論法では、この導出の段階で類似知識の適用を行う。すなわち、従来の導出が、同一の素式の肯定と否定を持つ二つの節から新しい節(導出節)を生成するのに対し、この二つの素式が類似していれば導出節を生成させる。具体的には、すべての導出において状況を考慮し、その状況  $V$  において、ゴールの否定(あるいはサブゴール)の  $\sim B$  と知識  $A \ B'$  ( $= \sim A \ B'$ ) の  $B'$  との類似度が、しきい値より大きければ、 $\sim A$  を生成させる。ここで、素式  $a$  と素式  $b$  の状況  $c$  における類似度を  $R(a,b|c)$  で表すとすると、この条件は  $R(B,B'|V) >$  となる。これは、 $A \ B'$  に類似する知識  $A \ B$  ( $= \sim A \ B$ ) を生成し、 $\sim B$  と  $\sim A \ B$  とで通常の導出を行うことと等価である。

この推論法を発展させ「言葉で考えるコンピュータ」の中核とする場合、類似知識の適用による探索空間の爆発が問題となるが、これについては、階層的な推論処理 [Tenenber89] の適用等により対処していくことを考えている。また、前節の言葉をベースとした知識表現形式を処理するには、推論メカニズム自体の改良も不可欠となる。例えば、従来の階述語論理では、引数は一般に個体を示す記号あるいは固有名詞が用いられるが、ここでは普通名詞の使用も認めている。この場合、その普通名詞で代表されるような個体を引数とする素式との関係が問題になる。例えば、「医者が患者を治す」を述語知識表現(ラベルは略)する場合、個体を意識すれば「治す( $x, y$ ) 医者( $x$ ) 患者( $y$ )」となり、普通名詞をそのまま用いれば「治す(医者, 患者)」となる。従って、単一化処理ではこれらが混在する場合への対処が必要となる。例えば、「治す(太郎, 花子)」と「治す(医者, 患者)」を単一化する場合、それまでに獲得された知識の中に「医者(太郎)」と「患者(花子)」があれば、これらは同一とみなす等の処理が必要となる。

#### 5. 自然言語文の述語知識への自動変換法

これまで、述部が動詞となる文(基本的に単文)を対象とした変換法について検討してきた [野中 03]。以下、この手法につ

いて述べる。この変換には、EDR 電子化辞書の日本語動詞共起パターン副辞書(以下、共起辞書)および単語辞書を用いる。まず、対象となる文を形態素解析した後、共起辞書から文中の動詞に関する全ての共起情報を取り出し、その情報(格助詞)をもとに深層格候補を抽出する。次に、それが意味的にその動詞の深層格になり得るかを単語辞書を用いてチェックする。最後に、全ての深層格が意味的に適切な共起情報を用いて述語論理(素式)に変換する(図2)。ただし、この共起辞書は複数の人により人手で作成されているため、深層格の決定に若干の揺らぎがある、適切な意味情報が付与されていない場合がある、収容されていない動詞がある等、完璧とはいえない。ここでは、これらに対処するため以下の処理を行っている。

深層格候補の単語の概念が、意味情報構成要素として示されている概念(識別子)の1つ上位の概念の下位概念となっていれば引数とする。

動詞が共起辞書に収容されていなかった場合や、取り出した共起情報の格パターンが一致しなかった場合には、IPAL 動詞辞書からその動詞の類語を抽出し、その共起情報を利用する。

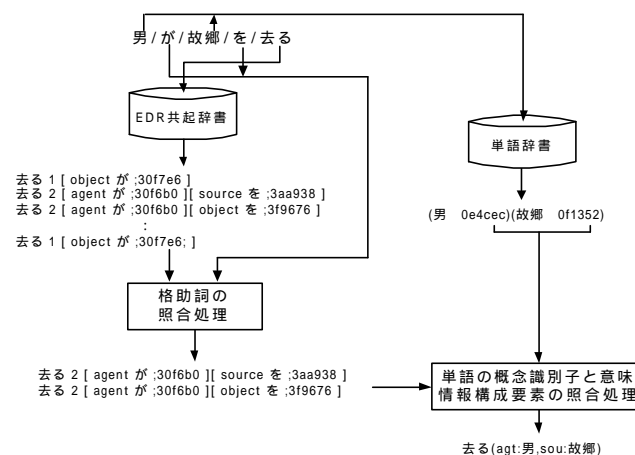


図2 述語知識変換の流れ

以上の方法について、変換対象の自然言語文としてライトハウス英和辞典[竹林 93]から100文、EDR共起辞書から200文、IPAL動詞辞書から100文を抽出し評価した。このうち、形態素解析が不正確、単語がEDR単語辞書に未登録等の場合を除く、329文を変換対象とした。評価結果を表3に示す。ほぼ90%近くが述語知識(素式)に変換できた。なお、変換できなかった主な原因は共起辞書に該当する格パターンが存在しなかったことによる。なお、以上の変換の逆変換に相当する、述語知識(素式)から自然言語文への変換法についてもあわせて検討した。これについてはほぼ100%正しく変換できた。

以上により自動変換の可能性について確認できたので、今後は、複文を含む3章で示した知識表現全般に対応した自動変換法の開発を進める予定である。

表3 評価結果

|                     | 変換率 |
|---------------------|-----|
| EDR共起辞書の共起情報のみを利用   | 71% |
| 概念識別子の適用範囲拡大(の方法)   | 13% |
| IPAL動詞辞書の類語を利用(の方法) | 4%  |
| 合計                  | 88% |

## 6. 常識知識ベースの構築

常識知識の獲得に関しては、これまで Lenat らの Cyc プロジェクトが百科事典からの獲得 [Guha90] を進めてきているが、最近ではインターネットを通して一般の人から獲得しようとする動き [Mindpixel] もある。

これに対し、我々は冒頭にも述べたように言葉の意味に関する常識知識のひとつとして概念ベースを構築し、従来のシソーラス以上の類似性判別能力を確認している [川島 03]。さらに、ごく一般的な事実知識の網羅的な獲得を目指し、国語辞書からの自動獲得について研究している。以下、これについて述べる。本方法は、国語辞書には少なくとも見出し語に関する事実知識が記載されており、かつその語義文には定型的な表現パターンが用いられていることに着目した方法である。具体的には、このパターンのうち最も多用されていると考えられる、「【見出し語】～を～する～。」というパターンの語義文を基本的な獲得対象とする。例えば、「【医者】患者を治療する人。」等である。即ち、この表現パターンは一般的に、「【A】BをCするD。」であり、A、B、Dは名詞、Cは動詞の語幹(サ変名詞を含む)である。また、DはAの上位概念となっている。従って、比較的簡単な構文解析を用いるだけで、「C(agt: A, obj: B)」の型の知識(先の例では、「治療する(agt: 医者, obj: 患者)」)を獲得できる。ただし、この基本的なパターンのみだと適用範囲が限定されるため、本手法ではいくつかの変形に対処可能としている[佐々木 03]。

本手法について評価するため、学研国語大辞典[金田一 88]中の主要語(4万語)の概念ベース[Nguyen 02]中の概念を選定)から人に関する常識知識の獲得を試みた。ここで人に関する見出し語は3,031個(多義を含めた語義文数は5,229個)あり、そのうち3,998個(76%)が動詞を含む語義文であり、多義の解消を行うと1,743個が本手法による知識獲得の対象となる語義文(広い意味でここで対象とした表現パターンに該当)となった。この1,743個に対して本手法を適用した結果、563個の知識が獲得された。また、そのうちの常識知識として妥当といえるものの割合は約78%であった。基本的には1つの表現パターンのみを対象とし、かつ簡単な構文解析しか行っていないこと等を考慮すると比較的良好な結果といえよう。今後、獲得率の向上と共に各種の表現パターンを対象とした知識獲得について検討していく予定である。

## 7. むすび

外界とのインターフェイスから内部的な処理まですべて言葉をベースとする「言葉で考えるコンピュータ」の構想、およびその基本部分についての検討状況について述べた。これまで得られている結果は、上記コンピュータ実現の第一歩とは言えるが、性能的にはまだまだ不十分であり、今後さらに検討を深めていく予定である。

## 参考文献

- [Nguyen02] Nguyen Viet Ha, 帆苺讓, 石川勉, 笠原要: 単語の意味の類似性判別のための大規模概念ベース, 情報処理学会論文誌, Vol.43, No.10, pp.3127-3136(2002)
- [Nguyen01] Nguyen Viet Ha, 石川勉, 阿部明典: 知識の類似性を利用した概略推論法, 電子情報通信学会論文誌, Vol.J84-D-I, No.4, pp.389-400(2001)

- [小林 01] 小林高志, Nguyen Viet Ha, 塩野好克, 石川勉: 感情表現と質問応答機能を備えた知的自由対話システム, 人工知能学会研究会資料, SIG-SLUD-A003-1, pp.1-8(2001)
- [野本 02] 野本和幸, 山田友幸: 言語哲学を学ぶ人のために, 世界思想社(2002)
- [兼岩 03] 兼岩憲, 佐藤健: DL: Description Logic, 人工知能学会誌, Vol.18, No.1, pp.73-82(2003)
- [Tenenbergs89] Tenenbergs, J.D.: Abstracting First-Order Theories, Change of Representation and Inductive Bias, pp.67-79 Kluwer Academic Publishers (1989)
- [野中 03] 野中昌行, 花田真理, 石川勉: 意味判定が容易な述語知識表現法とそれへの自然言語文の自動変換法, 情報処理学会, 第154回自然言語処理研究会, NO.154, pp.31-38(2003)
- [竹林 93] 竹林滋, 小島義郎: ライトハウス英和辞典: 研究社(1993)
- [Guha 90] Guha, R.V. and Lenat, D.B.: Cyc: A midterm report, AI Magazine, Vol.36, No.6, pp.32-59(1990)
- [Mindpixel] <http://www.mindpixel.com/>
- [川島 03] 川島貴広, 石川勉: 言葉の意味に関する類似性判別能力における概念ベースとシソーラスとの性能比較, 情報処理学会第65回全国大会, 2M-1(2003)
- [佐々木 03] 佐々木智彦, 益子公德, 野中昌行, 石川勉: 国語辞書からの常識知識の自動獲得法, 情報処理学会第65回全国大会, 2M-2(2003)
- [金田一 88] 金田一春彦, 池田弥三郎: 学研国語大辞典第二版, 学習研究社(1988)