

プログラム仕様に用いる 5L-1 自然語処理のための語句の意味定義

河野 俊二 御井 敬 菊野 亨 鳥居 宏次

大阪大学 基礎工学部

1. まえがき

ソフトウェア工学、及び、自然語処理の研究の進展に伴い、プログラム仕様を自然語で記述し、それを実行可能なプログラムへ変換する試みが多く報告されている^[4]。それに関連して著者らも適用分野をプログラム仕様限定して、自然語の意味を代数的な枠組みの中で定義する試みを行っている^{[2][3]}。

そこでは、プログラム仕様を記述している自然語の各文をある形式論理体系上の論理式に変換する。更に、論理式に現れる各語句の意味を代数的公理を用いて定義する。こうして求めた論理式と公理の集合でプログラム仕様の意味定義とする。本稿では、各語句の意味を定義する公理の記述について述べる。

2. 自然語処理システム^{[2][3]}

2.1 処理の概要

仕様の記述に用いる自然語(ここでは英語とする)を言語 L_{ns} と表す。いわゆる GPSG^[1] を用いて L_{ns} の構文則を与える。更に、 L_{ns} の各文を一階述語論理(L_1)と表す)上の論理式へ変換する規則(変換則)を定める。構文則と変換則を適用して求まる論理式の意味を L_{ns} の各文の意味と定める。

2.2 システム構成

処理の概要を図1の変換システムに示す。パーザは構文解析と一部の意味解析を行い、構文木(図2参照)を生成する。同時に、意味解析結果を表現する中間表現も生成する。図3に語句containsに対する中間表現の例を示す。中間表現の詳細については文献^[3]を参照されたい。

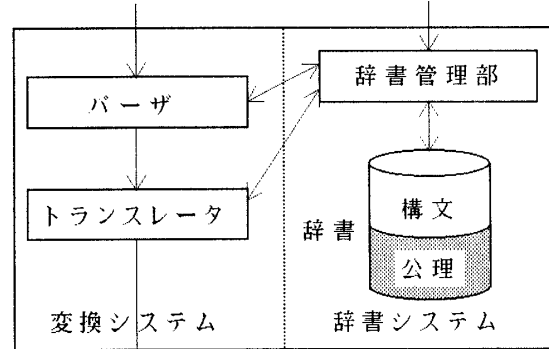
トランスレータでは変換則を適用して中間表現から論理式へ変換する。パーザとトランスレータにおいて利用する各語句の構文情報、中間表現、意味定義は辞書に記憶される。

2.3 辞書管理部

辞書は語句の構文情報と中間表現を記憶する構文辞書と、語句の意味定義である公理を記憶する公理辞書に分けられる。自然語の語句の中で、限定詞、接続詞、助動詞等は変換則の中に埋め込まれている。一般の動詞、名詞等の対象領域ごとに意味の定まる語句は辞書に記憶される。

辞書管理部が提供する主な機能を次に示す。

自然語 L_{ns} による仕様記述 語句の構文、意味定義



述語論理 L_1 上の論理式

図1 自然語処理システム

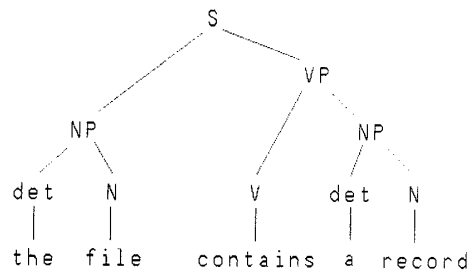


図2 構文解析木

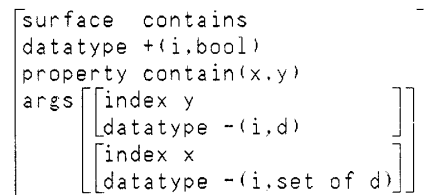


図3 中間表現

F1: 辞書への語句の登録、及び内容の更新の管理

F2: 語句の検索要求、定義要求の処理

ここで議論するのはF2中の定義要求の処理である。

3. 語句の意味定義

3.1 定義要求

自然語 L_{ns} の文の意味は一階述語論理 L_1 上の対応する論理式と、論理式中に現れる述語、関数、及び、データタイプの意味定義(公理)から構成される。ここでは次に示す3種類の意味定義の要求について議論する。

(1) データタイプに関する定義要求 (D, N)

D...Nがデータタイプであることを表す識別子
N...定義すべきデータタイプ名

(2) 述語に関する定義要求 (P, N, ω)

P...Nが述語であることを表す識別子
N...定義すべき述語名
 ω ...引数のデータタイプ名の系列

(3) 関数に関する定義要求 (F, N, ω , α)

F...Nが関数であることを表す識別子
N...定義すべき関数名
 ω ...引数のデータタイプ名の系列
 α ...関数が返す値のデータタイプ名

なお、上述の各要求はパーザから辞書管理部に対して出される。

3.2 基本方針

語句の定義要求が出されたとき、既にその語句の定義が辞書に存在する場合にはそれを利用する。新たに定義する場合でも、定義済みの述語、関数、データタイプや型構成子（集合、配列、レコード、…）を可能なら利用する。実際にどれを利用するかは選択は現段階では人手で行なっている。また、定義自身の記述も現段階ではすべて人手に頼っている。

4. 記述例

4.1 例1

次の英文を考えてみる。

The file contains a record.

この例文に対して変換則を適用すると、次の論理式

$\exists x[x \in \text{record} \wedge \text{file contains } x]$

が対応づけられる。ここで、 $e \in d$ は対象eのデータタイプがdであることを表す。これに伴って、3つの定義要求(D, record), (D, file), (P, contains, (file, record))が行われ、次に示す意味定義が求まる。但し、 $\forall x_1[x_1 \in d_1 \wedge \forall x_2[x_2 \in d_2 \wedge \dots \wedge \forall x_n[x_n \in d_n \wedge [\alpha]] \dots]]$ を $\forall x_1 \in d_1, \forall x_2 \in d_2, \dots, \forall x_n \in d_n[\alpha]$ と略記している。

1) containsの意味定義

$\text{contains}(x, y) \Rightarrow y \in x$

2) fileの意味定義

$\text{file} \Leftarrow \text{set of record}$

3) recordの意味定義

$\text{record} \Leftarrow \text{set of character}$

4) setに関する定義

$\forall d \Leftarrow \text{data type}, \forall e \in d$

[[set of d \Leftarrow data type]

$\wedge [\{\} \Leftarrow \text{set of d}]$

$\wedge [\forall st \Leftarrow \text{set of d} [\text{st} \cup \{e\} \Leftarrow \text{set of d}$

$\wedge \{e\} \cup \text{st} \Leftarrow \text{set of d}$

$\wedge \text{st} \cup \{e\} \Rightarrow \{e\} \cup \text{st}]]]$

$\forall d \Leftarrow \text{data type}, \forall st \Leftarrow \text{set of d}, \forall e \in d$

$[e \in \text{st} \Leftarrow [\exists st' \Leftarrow \text{set of d} [\text{st}' \cup \{e\} \Rightarrow \text{st}]]]$

4.2 例2

次の英文を考えてみる。

A stream of telegrams is available as a sequence of letters.

この例文に対して変換則を適用すると次の論理式が対応づけられる。

$\forall x[x \in \text{stream of telegram} \wedge \exists y[y \in \text{sequence of letter} \wedge [x \text{ is available as } y]]]$

従って、3つの定義要求(D, telegram), (D, letter), (P, is available as, (stream of telegram, sequence of letter))が行われ、次に示す意味定義が求まる。

1) is available asの意味定義

$[\forall x \in d, \exists y \in d' \text{ is_available_as}(x, y)]$
 $\Rightarrow [\forall x \in d, x \in d' \Rightarrow \text{true}]$

2) telegramの意味定義

$\text{telegram} \Leftarrow \text{sequence of letter}$

3) letterの意味定義

$\forall x \in \text{letter} [x \in \text{character}]$

4) sequenceに関する意味定義

$\forall d \Leftarrow \text{data type}, \forall e \in d$

[[sequence of d \Leftarrow data type]

$\wedge [e \in \text{sequence of d}]$

$\wedge [\forall sq \Leftarrow \text{sequence of d} [sq \cdot e \Leftarrow \text{sequence of d}]]]$

5) streamに関する意味定義

$\forall x \Leftarrow \text{stream of sequence of d} [x \Leftarrow \text{sequence of d}]$

5. むすび

現在、例題の解析を通して、語句の意味定義の蓄積を行っている。今後、自然語の語句の意味定義を支援するシステムの開発を進める予定である。特に、3.2で述べた定義の再利用、及び、人手によって記述された定義の構文的なチェック等の作業の軽減を図ることを考えている。

謝辞 日頃、ご指導を頂く大阪大学基礎工学部 高忠雄教授に深謝致します。また、辞書システムのインターフェース部の決定に関してご協力頂いた本学 関浩之博士、松村享氏に感謝致します。

参考文献

- [1] G. Gazdar, E. Klein, G. Pullum and I. Sag: "Generalized Phrase Structure Grammar", Basil Blackwell(1985).
- [2] 御井, 河野, 菊野, 鳥居: "プログラム仕様に用いる自然語のための辞書の設計", 昭和63年電子情報通信学会春季全国大会, D-372(昭63-3).
- [3] 並河, 松村, 関, 藤井, 嵩: "自然語によるプログラム仕様の, 抽象的データタイプを用いた構文及び意味解析", 情処研報, NL-65-5(1988-3).
- [4] 辻井, 上原: "ソフトウェア工学と自然言語処理", 情報処理, 28, 7, pp. 913-921(1987).