

自動理解を目的とする退院サマリの体系的分析†

横田 将生** 吉 武 春 光***
砂 川 賢 二** 中 村 元 臣**

現在までに、医療を目的とした実用的患者情報システムが数多く開発されてきている。その大部分は数値あるいは符号化されたデータを対象としており、自然言語で記述された文章データは取り扱っていない。しかしながら、文章データは患者情報のより大きな部分を占めており、治療や診断により重要と考えられている。最近、著者らは九州大学附属病院に保管されている退院サマリ（カルテの一種）中の自然言語文章データに関する自動理解処理の研究を開始した。本論文では、議論世界に関する知識に基づいたそれらのデータの体系的分析および自動処理の概要について述べる。

1. ま え が き

医療を目的とした患者情報システムの必要性が指摘され始めてから今日まで多くの実用システムが開発されてきている¹⁾。しかしながら、それらはほとんど検査値など一部の数値化された（あるいは、されやすい）具体的な情報についてのデータベースシステムであり、診断や治療に最も重要でかつ患者情報の大部分を占める自然言語で記述された抽象的な情報はエキスパートシステムにおけるルールのように表層的に用いられることはあっても、深層的な概念（意味内容）は取り扱われていない。その原因は自然言語の意味内容を取り扱う技術が十分でないという一般的な事情の他にそれらの自然言語データを含むカルテ類が書式のない手書で、しかも図や表が混在するという特殊な事情による。我々は長年、自然言語の意味記述およびその応用に関する研究を行ってきており、既に IMAGES-I, II, III と呼ぶ自然言語理解システム^{2)~6)}を作成している。本研究の目的はそのような経験に基づき自然言語医療データの意味内容を自動的に理解し医師たちの言葉による問合せに対応できる実用的な自然言語理解システムを作成することにある⁶⁾。

今回の研究対象は九州大学医学部附属病院循環器内科に保存されている退院サマリの文章および医師達はその退院サマリに関して発する質問文である。退院サマリは入院患者が退院する時に発行されるものであ

り、文章、図および表が混在しているのがわかる。記載されている文章は日本語が主体であるが、部分的あるいは挿入的に英語や特殊な記号による記述が出現することがある。現在、図表を除く手書の自然言語文章だけをできるだけ忠実にワードプロセッサを介して入力蓄積しているが、キーボードから入力できない特殊な記法や記号に関しては一意的に他の表現に変換して入力している。

以下では、計算機による退院サマリの自動理解処理の観点から、内容構成、言語（構文および意味）分析方法および結果、省略に関する考察、理解システム構成の概要について述べる。

2. 退院サマリの内容構成

対象としている退院サマリの記載内容は医療に関する専門知識によるものと患者の日常生活など一般常識によるものより成っており、議論される世界は極言すれば以下の三者関係となる。

- (A) 患者
- (B) 九大病院循環器内科
- (C) B以外の医療施設（または九大病院診療科）

退院サマリはこの三者が登場する物語とみなすことができ、次の三つの連続した事件がそのあらすじを構成している。

- (1) AがCを受診しBに紹介される。
- (2) AがBに入院し治療を受ける。
- (3) AがBを退院し報告がCになされる。

これらの三者が主な登場人物として物語が以下の七つの段落をもって展開される。

- (D1) 患者基本情報（患者氏名、性別、年齢、生年月日、退院番号、入退院月日、紹介医師）
- (D2) 退院時診断（診断名、病因学的所見、解剖学

† Systematic Analysis of Discharge Summaries for Automatic Understanding of Them by MASAO YOKOTA (Kyushu University Hospital), HARUMITSU YOSHITAKE (Division of Business Administration, Faculty of Commerce, Seinan Gakuin University), KENJI SUNAGAWA, and MOTOOMI NAKAMURA (Kyushu University Hospital).

**九州大学医学部附属病院

***西南学院大学商学部経営学科

*現在、福岡工業大学工学部言語情報工学研究所

(退院時診断)#1.Organic heart disease.
 etiological:unknown for AR.coronary atherosclerosis for OMI.
 anatomical:old myocardial infarction(anteroseptal,S54 onset).cardiomegaly(CTR82% on admission →76% at discharge).
 physiological:aortic regurgitation(III'),mitral regurgitation(I'),left ventricular failure(over on admission,latent at discharge).left ventricular hypertrophy,left axis deviation,atrial premature contractions.
 (病歴及び主要な入院時現症と検査成績)
 (CC)dyspnea,fatigue.
 (PI)リュウマチ熱の既往は明らかでない。生来健康であった。20年前、心臓弁膜症を指摘され治療を受けていた。S55年からは当科外来でもfollow upされていたが、CHFのため入院を繰り返していた。S58年12月頃より風邪気味となり、これを契機に夜間多尿、運動時息切れ、呼吸困難などが亢じた。本年1月10日の外来でCHFの治療が必要と判断され入院となった。
 (PH)S54 Pneumonia.
 (FH)heart disease(+),father:CVA,mother:myxoedema.
 (LH)Smoking(-).
 (PS)148cm,45kg,consciousness alert,orthopnea
 (意見)#1.今回の症状の出現時期とCTRの拡大が一致し、S58.12月では認められなかったAR murmurが出現していることから、AVのdysfunctionによるものと考えた。S52年(1内入院時)の時点でもLADの拡大がありMRを疑っていたようであるが、心カテではMRは認められていない。MRの出現はS52年以降ということになるが、今回のCHF症状はMRの関与もあるものと思われる。
 (入院中の経過)#1.dyspnea 入院後、塩分6g/日、drug freeで経過をみていたが、proxysmal nocturnal dyspnea,orthopnea出現しLasix 1Tを投与した。しかし、胸写上congestionの所見の改善なくmoist raleも聴取されたのでLasix 1.5T,Prazosin 3mgを投与して経過観察中である。exertional dyspneaが出現し始めた9月頃よりCTR拡大しており、症状が特に酷かった12月~1月頃CTR 60%と有意に拡大している。
 (退院時の治療・今後の方針)#1.ニトロ-β 6T 3n.d.E.,ニトロ-βR2T 1v.d.S.,ミニアレス 3.0mg 3n.d.E.

図 1 退院サマリ文の例

Fig. 1 Examples of discharge summary sentences.

の所見、生理学的所見、機能的所見)

- (D3) 病歴および主要な入院時現症と検査成績
- (D4) 入院中の経過
- (D5) 退院時の治療、今後の方針
- (D6) 意見
- (D7) 病院基本情報 (報告年月日、報告先、報告者)

病歴および主要な入院時現症と検査成績 (D3) はさらに以下の七つの小段落からなり、これらはほぼこの順序で出現する。

- (D31) 主訴
- (D32) 既往歴
- (D33) 生活歴
- (D34) 家族歴
- (D35) 現病歴
- (D36) 現症

(D37) 検査成績

上述の大段落のうち構文処理の対象になるのはD2~D6である。D1およびD7は定形化されており問題とはならない。D2~D6の大段落は活字印刷されたそれぞれの名称で導入される。またD3の小段落は手書された日本語、英語(またはその略)の名称で導入される。構造的に定形性が見られるものはD2だけである。D2はほとんど英語で書かれておりその構造はほぼ次の式(1)で表現できる。

(N D Etiological: E Anatomical:

A Physiological: P Functional: F)* (1)

ただし、N:番号、D:診断名、E:病因的所見、A:解剖学的所見、P:生理学的所見、F:機能的所見、()*:反復可。

残りの4段落は項目とデータの対の羅列、記号による現象の表現や英語表現を部分的あるいは挿入的に含むことがある日本語文章よりなっている。現在までに約40人分(約2千文章)の退院サマリを分析した。その結果、筆者と読者が暗黙裏に共通の議論世界を前提に情報伝達が行なわれるため、合理的(暗黙の知識または文脈情報で補充可能)な省略が多発しやすく軽度の文法(慣用)逸脱もしばしば起こっていることが明らかになった。例えば、主語はほとんど省略されている。図1にD2~D6における典型的な例を挙げている。

3. 言語分析および記述のための作業原理^{1)~7)}

本章では、我々の提案している人間の自然言語理解機構のモデルに基づき、自然言語医療情報に関する体系的言語分析・記述のための作業原理を明確にする。

3.1 自然言語理解機構のモデル

一般の自然言語体系は人間の日常的感覚知覚に基づく記号の体系である。そこで人間を外界(物理的世界)および内界(精神世界)を測定する装置と考えるならば、自然言語の担う概念(イメージ)はすべてこのような装置上の現象に還元されるはずである。この考えに基づき、我々はそのような測定器のモデルとして人間の生の感性経験記憶を座標とする抽象的空間(属性空間)を仮定し装置上での現象をそのような空間における軌跡と対応づけ、その軌跡を一定単位で分節し述語論理により計算可能な形式で記述する方法を考案した。この考え方は医療世界にも適用可能である。この場合の測定装置は人間(医師および患者)の感覚器官および各種の医療機器である。例えば、患者

の主訴（胸の痛みなど）は患者の痛覚で、医師の聴診は医師の聴覚を使って行われ、血圧や体重はそれ専用の測定装置が用いられる。したがって、このような測定装置に対応する属性空間を設定することになる。

医療情報世界の場合は自然言語体系を図2のような対応構造の中で取り扱う。図2において、 W_m は医療世界、 M_m は医療世界の内包的（イメージ）モデル、 D_m は M_m の論理的記述体系、および N_m は医療情報における自然言語体系をそれぞれ意味する。退院サマリの場合に当てはめて具体的に述べると、 W_m は2章での三者が主に関係する現実世界での事柄、 M_m はそのような事柄を測定器と対応する属性空間の現象として捉えたもの、 D_m は退院サマリの内容に計算可能な形式を与えるもの、および N_m は退院サマリに用いられている自然言語である。退院サマリの体系的言語分析・記述はこのような対応構造に基づいて行われる。

3.2 医療世界の枠組み

図2における M_m を体系的に設定するためには W_m を体系的にしき必要十分に把握する必要がある。図3は、このような観点から W_m の骨格をモデル化したものである。この図の意味するところは概略以下のようなものである。

「医師が患者およびその環境に対して作用（問診、検査、治療）し、その反応を種々の手段（人間の感覚器官や医用機器）によって発見し、事実としてあるいはそれらに知識（常識または専門知識）に基づく推論を施し判断（診断）として書類（カルテなど）に記述する。さらに、医師はこのような書類から患者の履歴に関する知識を得る。」

3.3 モデルに基づく作業手順

我々の自然言語解釈・表出方法では表層構造と表層依存構造の間、表層依存構造と意味構造の間において相互に変換が行われる。この後者の過程においては単語の意味記述（概念部および結合操作部）を介して相互変換が行われる。理想的な自然言語医療情報理解システムは図3における医師の「退院サマリ等の報告書類（カルテ）」解読能力および過程を

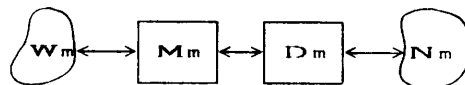
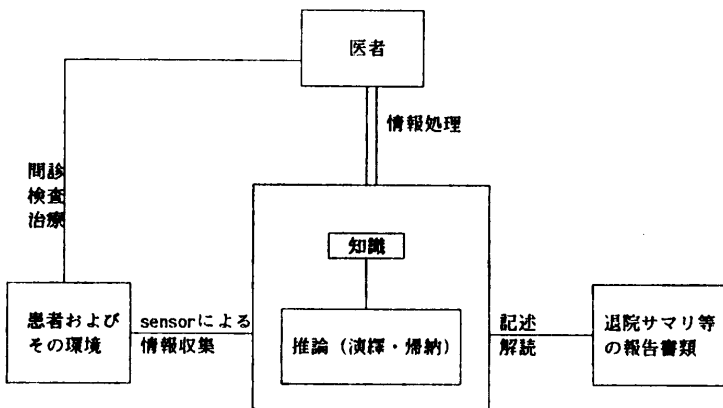


図2 医療世界における自然言語体系 ($W_m, M_m, D_m, N_m, \leftrightarrow$ はそれぞれ、医療世界、内包的モデル、論理的記述体系、自然言語体系、対応関係を意味する。)

Fig. 2 Natural language system in the world of medicine (where W_m, M_m, D_m, N_m , and \leftrightarrow mean the world of medicine, intensional model, logical description system, natural language system, and correspondence, respectively).



sensor: 医師の感覚器官あるいは機器

図3 医療世界の枠組み

Fig. 3 Framework of the world of medicine.

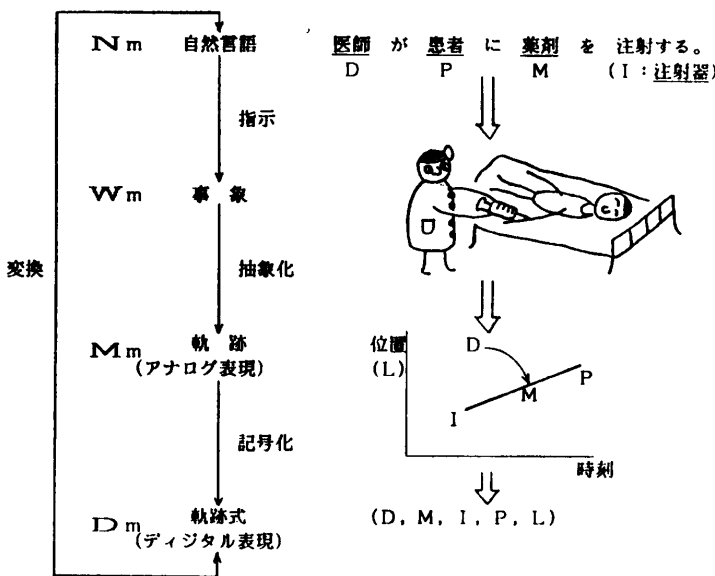


図4 医療世界における自然言語の体系的分析および記述

Fig. 4 Systematic analysis and description of natural language in the world of medicine.

模倣でき、その解読結果は医師の記述した内容と一致することが必要である。このような事柄を達成するために、我々はカルテに記載された文章の体系的言語分析・記述を図2のモデルに沿って、以下のような作業手順(P1-P5)で行っている。

P1) 退院サマリ文(Nm)とそれが指示する医療世界における事象(Wm)を対応づける。

P2) 事象と属性空間における軌跡(Mm)とを対応づける。

P3) 属性空間における軌跡を原子軌跡単位で分節し軌跡式(Dm)で表現する。

P4) 退院サマリ文の表層依存構造を生成する。

P5) 表層依存構造と軌跡式間の対応関係を分析し、単語の意味記述(概念部および結合操作部)を行う。

これらの作業のうち、P1-P3は人手で、残りP4、P5を計算機を用いて自動的あるいは会話型で行う。図4は具体例を用いて以上の作業手順を説明するものである。

4. 退院サマリの概念分析

自然言語が担う概念は事物概念(事象または物の概念)およびその補助概念に分類できる。事物概念は属性空間における軌跡と対応づけられる論理式で表現され、補助概念は属性空間の名称(体温など)あるいはそこにおける座標(38℃などの属性値)に対応し論理式を構成する項(term)に相当するものおよび論理的結合子(接続詞の概念など)に相当するものより成る。我々の理論では物の概念はそれが加担することのある事象概念によって規定される(例えば、「薬剤」

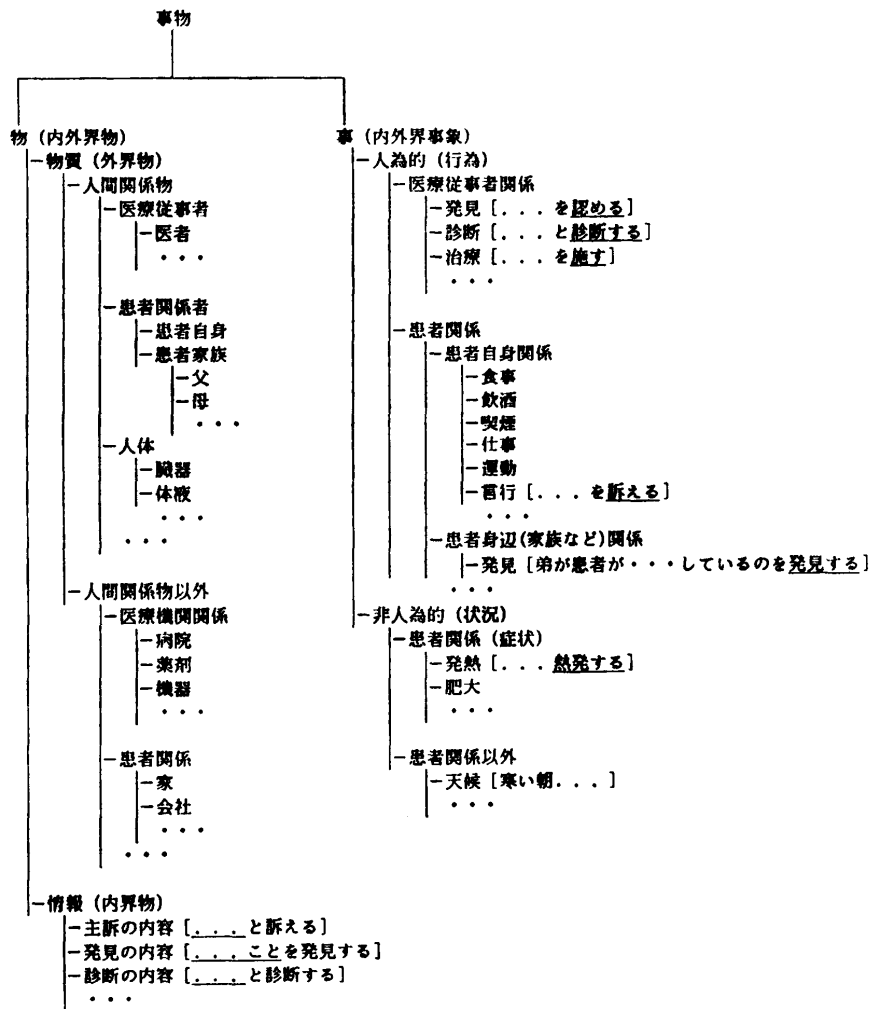


図5 退院サマリに出現する事物概念
Fig. 5 Entity concepts found in discharge summaries.

表 1 属性・事物の対および属性値表現の例
Table 1 Examples of attributes, objects or events, and attribute values.

属性	事物	属性値表現例
diagnosis	患者	低血圧
treatment	患者	入院加療
degree	事(症状)	強い, 症状に変化なかった
sequence	事	今回の発作
injection-mode	薬剤	NG 1T 舌下
finding	患者	heart murmur を指摘される
count (度数, 個数)	事物	～も, count-up, 3回, 3 Tables
color	物	淡い陰影
fever-pattern (熱型)	患者	熱型は変わらず
capability	事	日常生活ができるようになった
distance	移動行為	家から約 500m の距離を買物に行った
duration	症状	数分から 10 分程度
height (身長)	患者	Height 163.4 cm
weight (体重)	患者	weight 59.5 kg

は患者に投与される“物”のように)。したがって、自然言語概念のうちで最も重要なものは事象概念であり、それが記述される属性空間である。事象概念の分析は 3 章の作業手順のうち、主に P1-P3 が関係しており、その具体的作業は既に述べたとおりである。この章では、特に事物および属性に関する分析結果の概要を示す。

現時点において事象概念に関しては約 700 種類（動詞のみ）、物概念に関しては約 20 個の範疇（医療機関、医療従事者、患者、患者家族、生活環境、薬剤、体の部位、体液、排泄物、摂取物、治療機器、検査機器など）またそれらの事物が担う属性の種類に関しては約 70 個抽出している。表 1 は抽出した属性および属性値表現の一部を掲げたものである。事物概念は図 5 に示すように図 3 のモデルを頼りに top-down 的にも網羅的に整理することができる。

5. 退院サマリの構文分析

構文分析（形態素解析を含む）には 3 章の作業手順 P4 が関係している。人間による自然言語データの構文分析は使用する構文規則の種類や分析アルゴリズムまたはその適用方法の一貫性に問題があるので、その結果をそのまま計算機処理の場合に適用できることは経験上まれで、その都度部分的あるいは全面的見直しが必要である。このような点を考慮し、今回の退院サマリの構文分析では現在我々が保有している日本語構文解析システムを用いて行った。このシステムは日本語表層構造を表層依存構造に変換する機能を有し、Lisp で書かれた形態素処理部および Prolog で書かれた構文処理部よりなる。登録されている構文規則は 43 種類あり、分析作業の途中で規則の変更を行えば受理できるようなデータに遭遇したが、システムの能力は最後まで固定して行った。ただし、未知語の形態素情報および品詞情報はその都度与え蓄積した。

今回の構文分析の対象は退院サマリ 14 人分の D3（病歴および主要な入院時現症と検査成績）～D6（意見）に出現する日本語文（英語句の挿入を含む）379 文である。これらを我々の構文解析システムに入力して分析したところ、表 2 および表 3 に示すような結果が得られた。表 2 は各サマリごとに文章を以下の 3 群に分類整理したものである。

GROUP 1: 受理されたもの、

GROUP 2: 一般性のある構文処理規則を追加すれば受理可能なもの、

GROUP 3: 受理されるためにはそれぞれの意味内容に依存した例外的構文処理規則が必要であるもの。

表 3 は GROUP 2 または GROUP 3 に分類された文章の例とその理由を示したものである。これらの表から判断されるように、現状の構文解析システムでは退院サマリの文章の過半数が拒否されてしまうことになる。ところで、我々の理解システムにおいて構文処

表 2 退院サマリ文の構文処理難易度による分類*
Table 2 Classification of discharge summary sentences by difficulties in syntax analysis.

サマリ番号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	計
GROUP 1	9	9	12	13	14	10	14	10	10	7	13	10	10	8	149 (39%)
GROUP 2	11	6	14	7	18	17	7	30	12	9	5	13	7	7	163 (43%)
GROUP 3	2	3	4	7	7	10	5	8	2	1	0	5	7	6	67 (18%)
計	22	18	30	27	39	37	26	48	24	17	18	28	24	21	379 (100%)

* GROUP 1 は受理されたもの。GROUP 2, 3 は拒否されたもので原因解決の容易さの順序である。

表 3 GROPU 2, 3 の主要な原因とその例*

Table 3 Major factors of GROUP 2 and 3, and sample sentences.

- 1) 記号を用いた特殊な表現: [例 1] CTR の拡大 (47%→53%). [例 2] BP は 150/100 程度であり…[例 3] …を除いた状態—Herbesser 3T P.O 後 2 時間後—で…[例 4] RA の造影 (+).
- 2) 構文的補助語 (助詞など) の省略: [例 5] 九大入院し, …[例 6] AVR 施行する.
- 3) 不規則な羅列表現: [例 7] 発作時, 発作後 3 hr, 6 hr, 18 hr の GOT が (16→17→14→14) であり, … [例 8] seg 6 (46~33% diffuse lesion), seg 10 (71%), seg 9; 13 (65%), seg 11 (56%).
- 4) 文法逸脱: [例 9] RCA & LCX were almost intact であり…[例 10] ECG で I, aVL, V4~6 で T inversion 指摘され, …[例 11] 発作を, NG で経過を見た.

* 1)-2) は GROUP 2, 3)-4) は GROUP 3 の主たる原因である.

理結果 (表層依存構造) は意味構造に変換されるまでの過渡的存在である²⁾. したがって, 非文法的な表層構造から意味構造が直接得られるような意味内容に基づく推論処理が可能になるならば, 多数の例外的構文処理規則を採用することは避けられる. 我々は, このような推論処理を次章で述べるように省略概念推定処理の一部として考慮している.

6. 省略に関する考察³⁾

実験的な自然言語理解システムから実用的なものに至る過程には本質的に解決しなければならない多くの言語現象上の問題 (例えば, ambiguity, anomaly, vagueness など) が存在する. それらの中でも省略 (ellipsis) の問題は最も重要であると思われる. なぜならば, その他の問題は談話中に情報を追加することによって解消できる場合が多いからである. すなわち, それらは追加すべき (省略された) 情報の推定問題と考えられるのである. 我々は IMAGES-I, II, III で英語を対象とする自然言語理解の研究を進めてきたが, 英語等に比較すると, 日本語は構文的拘束が弱くその使用において語句の省略が頻繁に発生しやすい. そのため日本語談話処理では省略概念の推定処理は不可欠である. 本章では我々が既に提案している自然言語理解処理体系を基に日本語における省略に関して一般的に考察する. 省略概念の推定処理を考える場合, その推定範囲をどう絞るかが処理効率上大きな問題である. したがって, このような考察には大いに意義が

ある.

6.1 省略の定義

自然言語理解システム IMAGES-III は入力された自然語談話に文単位で順次構文処理および意味処理を施し, 最終的に得られた無矛盾な談話意味構造に関して質問応答を行うことができる. 図 6 は IMAGES-III における文単位処理を体系的に示したものである. このような処理の結果, システムが受理するのは文脈あるいは知識との比較により「真」あるいは「疑 (真偽不明)」 (これは新事実と認識する) と判定されたものだけである. ところで, 既述のとおり, 退院サマリに出現する文の過半数は意味理解処理に至るまでにシステムに拒否される運命にある. カルテ等の性質からするとその内容は有意義で無矛盾と考えるのが妥当である. そこで, システムが受理できないのは「入力文を有意義で無矛盾とする情報が省略されている」と考えるのは極めて合理的である.

自然言語の意味理解という立場では, 「曖昧さ」などと同様に「省略」は表出側 (話者) よりも解釈側 (聞き手) の問題である. すなわち解釈側がどのような意味処理体系で対処するか (理解するために何を必須の情報としているか) によってその性質あるいは定義が異なる. ところで, 談話を文が接続詞 (これは省略される場合の方が多い) で結合されたものと定義すると重要なのは一文中における省略である. 我々の属性空間モデルに基づく自然言語の意味記述は概念部および結合操作部より成っている. そこで, それらの意味記述に依拠する意味処理体系での一文中における「省略」を一般的に次のように定義する.

(定義 1) ある文の表層依存構造を意味構造に変換する際, 単語の結合操作部に記述されている操作命令が一つでも使用されなかった場合, その文には省略がある.

退院サマリのように議論世界が極度に限定され, し

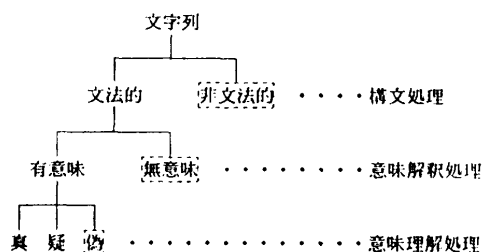


図 6 IMAGES-III における理解処理の概要
Fig. 6 Outline of understanding process of IMAGES-III.

かも談話の有意義性および無矛盾性が保証されている場合には、さらに次の定義を追加することも可能である。この定義の場合は、文の文法性（表層依存構造の存在）をも前提としないため、上記の保証がなければ省略概念の推定は困難である。

（定義2）ある文を意味構造に変換する際、単語の結合操作部だけではすべての単語概念がその議論世界で“真”または“疑（真偽不明）”となる唯一の文意味構造に統合できない場合、その文には省略がある。

6.2 省略の分類

図7は自然言語理解システムIMAGES-IIIの意味理解処理に基づいて省略の種類およびその例を求めたものである。（定義1）に合致するのはタイプdの場合だけで残りはすべて（定義2）の範疇である。その他、名詞の接続（例1）などや名詞の格助詞「の」による接続（例2）なども本質的には省略と考えられる。このような省略に対してシステムは第一に文脈情報、第二に事前知識（議論世界の知識や常識など）により省略概念を推定することになる。

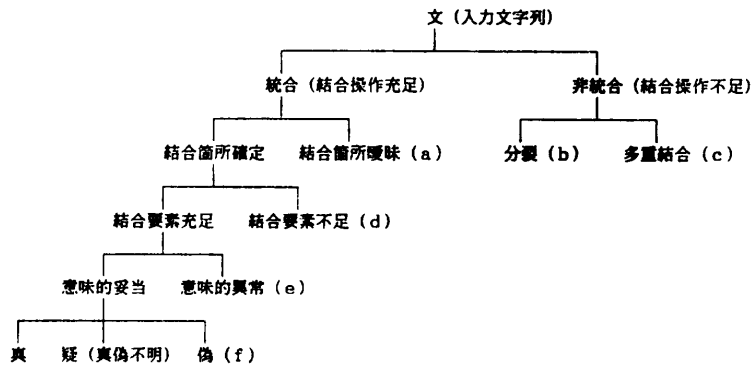
（例1）近医受診（＝近医を受診すること）

（例2）患者の会社（＝患者が勤めている会社）

また、未知語に関しては、それを含む表層依存構造あるいは意味構造が既知語のみから生成される構造から孤立して存在することになり、タイプbの省略として認識される。その品詞や意味などはシステム自身が既に所持している言語や議論世界に関する知識から推定するか、人間から学習させられることになる。

7. システム構想

IMAGES-IIIは、議論世界が限定されない一般分野の談話文章を対象に、構文処理および意味処理を順次独立に行いながら単語概念から句、節、文そして談話の概念を生成していく上昇式（bottom-up）自然言語理解システムである。ところで、現在我々が対象としている循環器内科の退院サマリでは、議論世界がかなり明確に限定されており省略（6章の「定義2」の意味）が多発するため、構文処理と意味処理を融合し、



- a. 助詞などの省略
（例）近医, 患者, 受診する。 → 近医を患者が受診する。
- b. 述語などの省略
（例）午前中, ミニプレス 3 mg。 → (医者が患者に) 午前中, ミニプレスを 3.0 mg 投与した。
- c. 述語などの省略
（例）発作をニトログリセリンで経過を見た。 → (医者が) 発作をニトログリセリンで対処し経過を見た。
- d. 用言に対する必須要素などの省略
（例）100 mg の食塩を摂取した。 → 患者が 100 mg の食塩を摂取した。
- e. 何らかの必須概念の省略（欠落）
（例）患者の血圧が大きい日は… → 患者の血圧 (の) 変動が大きい日は…
- f. 文脈の展開における飛躍（新事実記載漏）か、事前知識更新の省略

図7 意味理解過程において弁別される省略の種類およびその例
Fig. 7 Types and examples of ellipsis discriminated through the understanding process.

しかも話題を予測しながら、すなわち下降式（top-down）の処理を同時に進めていくことが必要である。そのためには、まず退院サマリの議論世界を事前に把握し理解システムにおける下降式処理に必要な十分な予備知識を抽出し、構文および意味情報を同時に参照できる処理系の開発が必要である。第3章において、一般的な医療世界の骨格的モデルとして図3を提示した。既に述べたように、対象としている退院サマリは七つの主題（段落）から成っている。それぞれの主題を構成する文の話題は図3のモデルに従うと図8のような生成規則で捉えられる。これらの規則における終

- （退院サマリ文）→（事実）|（判断）
- （事実）→（問診）|（検査）|（治療）
- （判断）→（演繹推論結果）|（帰納推論結果）
- （問診）→（主訴）|（病歴）|（家族歴）|…
- （検査）→（心電図）|（超音波心臓図）|…
- （治療）→（投薬）|（手術）|…
- ……

図8 生成規則によって範疇化された退院サマリ文の内容
Fig. 8 Contents of discharge summary sentences categorized by production rules.

端記号は「事実」に関しては人間の感覚器官や循環器内科で使用される医用機器に、「判断」に関しては循環器内科医の判断行為に対応する属性空間の事象を指示することになる。

図9は我々が意図している自然言語理解システムの概念図である。省略や文法逸脱を含む退院サマリのような文書を計算機に理解させるには、まず議論世界に関する知識を計算機に記憶させ、その知識と文脈情報を利用した意味情報優位型の文章解析を行わせる必要がある。したがって退院サマリ文の解析処理は形態素、構文、意味解釈処理を統合して「必ず有意義で無矛盾な情報が表現されている。」という前提で行う。例えば、一見非文法的な(例3)*の場合、議論世界の知識から主に次の二つの事実を導出し陽に文の意味として記述することになる。

(例3) 発作をニトログリセリンで経過を見た。

(事実1) “患者の(容体の)「経過」を「見る」”

(事実2) “「ニトログリセリン」は狭心症の発作に対して用いられる”

この例文の処理手順は概略次のようになる：

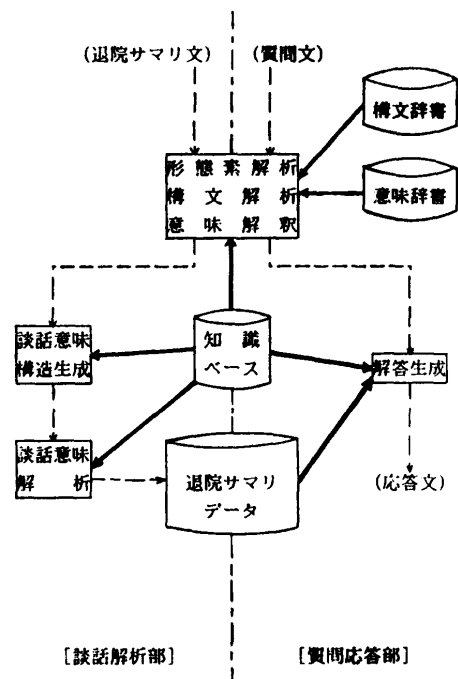


図9 退院サマリ理解システムの概念図

Fig. 9 Configuration of discharge summary understanding system.

* この場合の格文法的観点からの非文法性は、語句の省略に起因すると考えられる。すなわち、「で」の後にたとえ「対処し」を補うと文法的になる。

1) 意味辞書(常識)および知識ベース(専門知識)を参照することにより“経過を”および“見た”の間の表層依存関係が概念的に検査され、(事実1)に基づき受理され、概念構造(M1)に変換される。

2) 意味辞書および知識ベースを参照することにより“ニトログリセリンで”および“見た”の間の表層依存関係が概念的に検査される。意味辞書に基づき格助詞“で”と結合しうる“ニトログリセリン”の「用途(事実2)」が知識ベースより抽出され、概念構造(M2)が生成される。

3) 意味辞書および知識ベースを参照することにより“発作を”および“見た”の間の表層依存関係が概念的に検査され、“経過を”との重複結合を理由にM1との統合が拒否され、タイプCの省略が発生していることが認識される。

4) 既に生成されている文脈概念のうち、同一文内のM2がまず選択されて“発作を”の概念構造との統合が試みられ、「発作」を共有概念として成功し、概念構造(M3)が生成される。[この場合M3はM2と同一である。]

5) 概念構造M1およびM3を結合し、例3全体の概念構造M4を生成する。[M4を言葉で表現するとおおよそ「(主治医が)患者の狭心症の発作にニトログリセリンを用い、その経過を見た」となる。]

図10は医師が退院サマリからどのような情報を得ようとしているかを実際に調査し自然言語質問文として収集した167例中の一部である。これらの例から推察されるように単なる事実検索だけに留まらず統計処理等を含む高度な処理を施した二次的な情報を期待している。以上の事実より構築を予定している実用システムは医学知識や一般常識の他に統計学などの知識を援用する高度な自然言語処理機能、さらには自動プロ

1. 当科に入院しAMIで死亡した患者のうち、発病よりX日目の死亡数及び死亡率を教えてください。
2. AMIで死亡した患者のうち、死亡当日もしくは病状急変時にECG上ST上昇が見られた患者は何人いますか？ 又、その名前と退院番号を教えてください。
3. 当科入院患者のうち、診断として最も多い疾患は何ですか？
4. Dopplerにて指摘された逆流のfalse positiveはどれくらいありますか？
5. 当科に最も多くの患者を紹介して頂いた病院はどこですか？

図10 医師の発した質問の例

Fig. 10 Examples of physicians' questions.

グラミング機能を有しなければならないことがわかる。現時点の目標は字面の一致による検索から概念による検索を念頭に置いている。すなわち、表現の違いに余り影響されない情報検索を考えている。

8. む す び

退院サマリの体系的分析と理解処理の概要について述べた。特に理解処理に関してはこの分野に特有の省略や文法逸脱のほか、名詞の連接や格助詞「の」の問題など一般的な日本語処理の問題が多くある。これらは未知語に関する問題を含めて実用的な自然言語理解システムを作成する過程で必ず遭遇し解決しなければならないものであり、省略概念推定処理の延長上で解決を図っていかねばならないと考えている。このような処理を実現するには、形態素処理、構文処理および意味処理を効率的に融合し、処理の上昇式と下降式の切り換えを円滑に制御する必要がある。例えば、上昇式の処理中に鍵となる主要な事物概念を入力文中に検出したら、その概念が関連する話題に焦点を合わせ下降式の処理に移行するなどの方策が考えられる。現在、このような効率的な処理方法を人間の思考過程に求めるためおよび方法論の妥当性を検証するために、思考心理実験にも着手している⁹⁾。今後はこのような心理実験、大量の退院サマリおよびユーザが発する質問文の分析、およびプロトタイプシステム¹⁰⁾の改良を継続することにより目標を達成したい。また、その間得られるまとまった知見に関してはその都度報告する予定である。

なお、本研究の一部は文部省科学研究費（一般研究 C62580024、特定研究 62210010）によっている。

参 考 文 献

- 1) 吉武, 横田, 田町: 自然言語理解システム IMAGES-I の構文解析過程について, 電子通信学会論文誌, Vol. J 67 D, No. 10, pp. 1147-1154 (1984).
- 2) 横田, 吉武, 田町: 自然言語理解システム IMAGES-I の意味解釈過程について, 電子通信学会論文誌, Vol. J 69 D, No. 3, pp. 777-784 (1986).
- 3) 横田, 吉武, 田町: 自然言語理解システム IMAGES-I の出力合成過程について, 電子情報通信学会論文誌, Vol. J 70 D, No. 11, pp. 2267-2272 (1987).
- 4) 横田, 吉武, 田町: 中間言語による自然言語理解の試み, 情報処理学会「知識工学と人工知能」研究会資料, 31-4, pp. 1-8 (1983).
- 5) 吉武, 横田, 河口, 田町: 自然言語理解システム IMAGES-III, 知識情報処理 シンポジウム, pp. 17-26 (1985).
- 6) 横田将生: 高度利用を目的とする自然言語医療情報の体系的分析, 「1988 情報学シンポジウム」, pp. 89-100 (1988).
- 7) 横田, 河口, 谷口, 田町: 凶形・言語世界における知識の表現と処理, 信学会研資, AL 80-55, pp. 11-20 (1980).
- 8) 横田将生: 日本語談話における省略の認定, 「人工知能システムの枠組み」シンポジウム, pp. 95-100 (1987).
- 9) 横田将生: 自然言語理解に関する心理実験について, 第2回情報処理学会九州支部研究会資料, pp. 133-141 (1988).
- 10) 横田, 吉武ほか: 退院サマリの理解処理 (第1~3報), 第34, 35, 36回情報処理学会全国大会論文集 (1987-1988).
- 11) 例えば, Proc. of MEDINFO '86; 第7回医療情報学連合大会論文集 (1987).
(昭和63年3月11日受付)
(昭和63年10月7日採録)

横田 将生 (正会員)

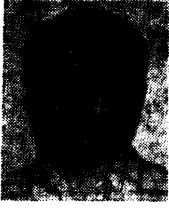


昭和47年九州工業大学工学部電子卒業。昭和52年九州大学大学院博士課程修了。同年同大工学部助手。昭和54年同大医学部附属病院講師。昭和63年福岡工業大学工学部電子工学科教授。言語および凶形の意味理解、医療・生体情報処理の研究に従事。工学博士。電子情報通信学会、人工知能学会、認知科学会各会員。

吉武 春光 (正会員)



昭和55年九州大学工学部情報卒業。昭和60年同大大学院総合理工学研究科情報システム学専攻博士課程修了。同年九州大学工学部情報助手。昭和61年西南学院大学商学部経営常勤講師。昭和62年助教授。自然言語理解および機械翻訳の研究に従事。工学博士。電子情報通信学会、人工知能学会各会員。

**砂川 賢二**

昭和49年九州大学医学部卒業後、同大学にて臨床研修を受けるとともに、心力学の研究に従事。昭和53年に米国マリーランド州のジョンスホプキンス大学の生物工学部および医学部に留学。講師・助教授を歴任し心力学・システム工学の教育・研究にあたる。昭和58年に九州大学医学部循環器内科の助手として帰国し臨床、教育およびシステム工学や心力学の研究に従事している。日本内科学会、日本循環器学会、日本ME学会、アメリカ心臓病学会各会員。

**中村 元臣**

昭和2年3月5日生。昭和25年九州大学医学部医学科卒業。医学博士。九州大学医学部・心研、循環器内科勤務。主たる研究テーマは虚血性心疾患の発生機序。著書：内科，狭心症，心筋虚血障害，冠攣縮。