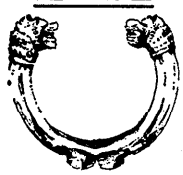
Vol. 34 No. 11　　情　報　処　理　　Nov. 1993

連載講座



自然言語処理入門—I

現状と歴史を概観しよう†

岡田直之††　中村順一††

次のような方はいらっしゃいませんか？

＊　若手技術者，研究者で，自然言語処理をこれから勉強しようとしておられる方，

＊　若手技術者，研究者で，自然言語処理について専門書の理解や実験に取り組んではいるが，いま一つポイントのつかめない方，

＊　熟年技術者，研究者で，自然言語処理の世界をおおざっぱに知っておきたい方，

＊　理工系大学の下級生，高専の上級生，短大生，専門学校生で，計算機を少し勉強し，自然言語処理に興味をもっておられる方，

＊　文系の学生，研究者で，計算機を少し勉強し，自然言語処理に興味をもっておられる方，

＊　その他，計算機を少し勉強し，自然言語処理に興味をもっておられる方．

もしこのような方がいらっしゃるなら，この初級講座をお勧めしたい．5回にわたるこの講座を読み通すと，自然言語処理が一とおり理解でき，また読者自身である程度処理システムを作ることができるようにと，筆者らを始め講座の関係者は意図している．図-1をご覧願いたい．情報系学部の3年生が本講座と同じ程度の学習によって作成した処理システムの実験例である．“Taro drives a sports car which sometimes breakes down on the street” という英文の構文構造および意味構造が解析され，その結果がそれぞれ“木構造”および“リスト構造”で示されている．初めての方にいきなり核心部分を十分な説明もなく紹介して恐縮であるが，本講座を通じてこのようなことを理解し，自分自身でシステム作りをしていただきたいと考えている．

この講座では，格別に難しい数学とか，複雑な計算機の構造とか，あるいは高度なプログラミングの技術とかを前提にはしていない．高校程度の数学や英語の常識とプログラムの初歩的知識があれば，自然言語処理を理解できるように企画されている．また，イメージの湧きにくい抽象的な表現や論文調のあまりに簡潔な表現は避け，できる限り具体的で丁寧な説明を試みている．さらに，理論や技法を天下り的に示すことは避け，それらの根拠あるいは背景を明確にした上で，どのようにして導かれたかを示そうとしている．読者自らが考え，発想を豊かにし，そしてシステムを構築する力をつけていただこうと考えている．

1. はじめに

本章では，自然言語処理の現状と歴史について述べる．初めに，自然言語処理の目標を示した上で，その実現に向けて現在行われている取組みならびに解決しなければならない課題について述べる．機械翻訳の実例にも少し触れよう．

次に，自然言語処理が今日に至った歴史を振り返る．機械翻訳を軸とする実用化志向の言語処理，知能の解明とその応用を目指す人工知能志向の言語処理，そしてそれらを支える言語理論について述べ，将来に向けての参考としよう．

最後に，本講座の構成について述べる．

1.1 自然言語処理の現状

1.1.1 目標と取組み

自然言語処理 (Natural Language Processing*) とは，プログラミング言語のような人工の言語に対し，日本語とか英語，ロシア語といった，人が日常話したり書いたりする言語を計算機で処理す

† Outline of Natural Language Processing: Current State and History by Naoyuki OKADA and Jun-ichi NAKAMURA (Department of Artificial Intelligence, Kyushu Institute of Technology).

†† 九州工業大学情報工学部

＊ 本講座では，しばしば使用する用語を簡単のため英語名の頭文字で略記することがある．たとえば“自然言語処理”はNLPと略記する．そこで，以下において英語名の頭文字に下線が施されている場合は，ご注意願いたい．

```
SENTENCE
        Taro drives a sports car which sometimes breaks down on the street.

STRUCTURE                    TOTAL 12

No. 3     time : 78 msec
  | -sentence
     |-sent
     |   |-s
     |       |-np
     |       |   |-nou 2 ---- taro
     |       |-vp
     |           |-vrb 1
     |           |   |-drive
     |           |   |-suffix ---- es
     |           |-np
     |               |-np
     |               |   |-det 1 ---- a
     |               |   |-np
     |               |       |-ajp
     |               |       |   |-adj ---- sports
     |               |       |-np
     |               |           |-nou 1 -------- car
     |               |-rpn ---- which
     |               |-sent
     |                   |-s
     |                       |-np ---- trace
     |                       |-vp ---- trace
     |                           |-vp
     |                           |   |-vrb 1
     |                           |       |-adp
     |                           |       |   |-adv 1 ---- sometimes
     |                           |       |-vrb 1
     |                           |           |-break
     |                           |           |-suffix ---- es
     |                           |           |-down
     |                           |-prp
     |                               |-pr ---- on
     |                               |-np
     |                                   |-det 2 ---- the
     |                                   |-np
     |                                       |-nou 1 ---- street
     |-prd ---- .
```

注 1. 与えられた文に対して TOTAL 12 の構造解析結果が得られ，No. 3 の解析結果が示されている．
2. S, np, vp … は，それぞれ文，名詞句，動詞句，…といった，文の構成要素を意味する記号．
3. nou 2, verb 1, det 1, … は，それぞれ名詞 2，動詞 1，限定詞 1，…といった，品詞を意味する記号．
4. たとえば "S|-np" という構造は，S が np と vp から構成されていることを示す．

```
S|-np
 |-vp
```

（a） 構文構造

図-1 文の構文および意味構造の解析

ることを指す．したがって NLP の目標は，そのような話し／書き言葉を利用者の要求に沿って処理することにある．利用者の要求はいろいろあるが，以下では，社会的ならびに学術的な面からいくつかの取組みを紹介しよう．

まずあげられるのは，**ヒューマンインタフェース** (human interface) あるいは **自然言語インタフェース** (natural language interface) とよばれる部門である．多くの人々が計算機に接する機会が増えている．計算機が出現した当初は，計算機の構造やプログラムを知っている専門家だけしか操作ができなかった．しかし最近では計算機，特に端末の操作もずいぶん改善され，素人でもかなりの程度用意されたプログラムを利用できるようになってきた．人と計算機が情報を交換するとき，人の側にとって好都合なメディアは，やはり言葉や図形である．単純な指示を与えるならたとえばメニュのような方式やカーソルのような道具立てでも可能であろうが，複雑な指示や微妙な説明になると，やはり言葉でこれこれしかじか，と伝え

```
[drive,
 [evt, [operation], [sfrm/pl_3]],
 [taro, [agt, [human], sbj]],
 [car,
  [obj, [movable＊thing], obj1],
  [sports, [att, [sports], mdf], [car, [pos, [thing], sbj]]],
  [a, det1],
  [break_down,
   [att, [miscellaneous], mdf],
   [evt,
    [miscellaneous],
    [sfrm/pl_1],
    [form/rpn]
   ],
   [car,
    [obj, [thing], sbj],
    [sports, [att, [sports], mdf], [car, [pos, [thing], sbj]]],
    [a, det1]
   ],
   [sometimes, [att, [time], mdf]],
   [street, [place_at, [thing], mdf], [the, det2]]
  ]
 ]
]
```

解析結果の意味は，概略次のとおり．
1行目：drive が文の意味全体を支配している．
2 〃 ：事象 (evt) としての意味内容は，operation.
3 〃 ：taro は，〔drive の動作主 (agt) で，意味内容は human，そして文中，主語 (sbj) となる〕.
4,5〃 ：car は，〔drive の対象 (obj) で，意味内容は movable かつ thing，そして文中，目的語 (obj1) となる〕.
6 〃 ：sports は，〔属性 (att) で，意味内容は sports，そして文中，修飾子 (mdf) となり〕，かつ〔car は，属性 (att) を所有するもの (pos) で，sports の主語 (sbj) となる〕.

(b) 意味構造

図-1 文の構文および意味構造の解析

ることができれば，これに越したことはない．また言葉で言い表しにくい部分を図を描いて説明できれば，さらに好都合である．どのような言葉や図形でも計算機がそれを理解して処理してくれる，というのは，究極に近いかなり将来の目標である．現在は，言葉にしろ画像にしろ，ある特定の分野に的を絞って計算機と対話を行うシステムの実現を目指している．

もう一つは，国際交流の部門である．日本は島国で，ややもすれば閉じた世界でものごとを考える傾向にあった．昨今，観光，留学，企業活動，学術会議など，急速に世界的規模で行動，思考する機会が増えてきた．ところが世界には多くの言語があり，母国語以外の言語を習得することは，一般に容易なことではない．このことが国際交流の大きな障害になっている．これはひとり，日本だけの問題ではない．ご承知のようにヨーロッパを始め，世界の多くの地域で言語の障害が種々の問題を引き起こしている．もし他国語で書かれた文書を母国語に翻訳してくれるパソコンや，マイクに向かって話すと即座に目的の言語でスピーカから発話する携帯用通訳器が出現すれば，前述の問題は相当改善されるに違いない．もちろん現在でも，翻訳用のパソコンや携帯用翻訳器が市販されてはいる．しかし機能的にもまだ十分成熟しておらず，改良に向けて研究者や技術者たちが熱心に取り組んでいる最中である．また最近では音声処理を含めた自動通訳システムが研究され，限られた範囲を対象とするなら実用化も期待されている．

さらに，**オフィスオートメーション**とよばれる部門がある．事務部門の合理化は，まず文書処理から始まる．会議の案内状とか品物の発注書とか，いわゆる紋切り型の文書を，毎回手書きするのは厄介である．また長い原稿を推敲する場合もそうだ．わが国では，片かな，平がな，漢字と文字の種類が多く，しかも多数の漢字を用いる事情もあって，欧米のようなタイプライタがあまりなじまなかった．しかし**ワードプロセッサ** (word

processor), いわゆる“ワープロ”の出現がこの問題を大きく改善させ，通常のタイプライタ以上に重要な機能を発揮することになった．現在は，文書の検索や要約あるいは画像データとの結合など，さらに高度な文書処理への取組みがなされている．

以上は，社会上あるいは実用上の取組みといえよう．これとは別に，学術上の取組みもある．実をいうと，人が言葉を理解したり発話したりするメカニズムはまだ十分解明されていない．言語活動は心理活動全般と密接に結びついており，NLP の研究には，言語学，心理学，あるいは生理学などの分野が深く関わってくる．これらの人々が工学の人たちと一緒になって，計算機を使って知能の解明をしようとするのが，**認知科学** (cognitive science) あるいは**人工知能** (Artificial Intelligence) とよばれる分野である．計算機にいろいろな知識や推論の規則を組み込み，繰り返し実験して，逆に人の知的振舞いを推測しようとするのである．もし知能のメカニズムがより詳細に解明されるなら，エキスパートシステムを始め現在の AI システムはずいぶん改善されることであろう．また，たとえば記憶の構造を参照して効率の良い語学の教育方法など，人文分野への応用も期待されよう．さらに将来，もし怪我などで言語障害を起こした人の神経回路に，言語処理用のチップを埋め込んで機能回復ができるようになるならば，医療面で大きく貢献できることも夢ではなかろう．

1.1.2 課　題

NLP は 1940 年代から取組みがなされ，技術的にも大きく進歩したにもかかわらず，人と同じように柔軟な対話のできるシステムは，現時点では見当たらない．そこで，これからの NLP が取り組まなければならない課題について，二，三述べよう．

まず初めにあげられる点は，対象としての言語の性質をより深く解明する，ということである．比較のため，国際線の飛行機の座席予約システムを考えてみよう．世界的規模で支店が展開し，大規模で複雑なシステムである．しかしながら，座席の予約，管理，変更など，処理すべき基本的項目や方式は定まっている．なぜなら，複雑ではあっても，すべて人が人工的に約束したものだからである．それに対し自然言語は，長い歴史的過程を経て自然発生的に創られたものである．座席予約システムほどきちんとした取り決めとか計算方式があるわけではない．**図-2** の例を考えてみよう．同じ“名詞＋の＋名詞＋の＋名詞”の並びであるが，第1の“名詞＋の”が(a)では第2の“名詞＋の”を，(b) では第3の名詞をそれぞれ修飾している．人は直感的にこの修飾関係を識別しているが，これを計算機に識別させることはやさしいことではない．なぜなら，そのためには，多分，物品の所有とか季節による衣替え，といった知識が関係してくると推測できる．所有関係や衣替えに関する知識だけならまだしも，この種の識別全体に必要な知識を明確にし，さらに計算機で処理できる形で示すことは，困難な作業といえよう．**1.2.2** では，これに対する研究者や技術者の取組みを紹介し，今後に向けての参考としよう．

(a)　(b)

図-2　名詞の修飾関係

もう一つは，大規模なデータを蓄積する点である．人に近い翻訳を目指すと，単語辞書として 10 万語ないし 20 万語くらいの大きさが必要といわれている．これを記憶させるハードウェアはあまり問題ない．問題なのは，その中の文法的ならびに意味的な種々のデータを，どのように表現し，蓄積するかというソフトウェアもしくはデータ工学である．自然言語の性質そのものがよく分かっていないからには，それを調べることから始めなければならない．**1.2.1** で，これまでに傾けられてきた努力を概観しよう．

3番目の点は，柔軟性に富んだシステムを実現する，ということである．銀行のオンラインシステムにしろ，乗り物の座席予約システムにしろ，一度システムを作り上げてから，個々のお客ごとに特別の利息計算をしたり，特別の予約を受けつけたりすることはあまりないと思う．ところが言葉は生きているので，用法が変化したり，人によって表現が変わったりする．人の場合は，少々間違った表現でも意味理解できる．計算機の場合は，少し間違ったり，定めた以外の用い方をすると，対処できないことが多々ある．人に近い，柔

原文

Historically, these grammar formalisms are the result of separate research in computational linguistics, formal linguistics, and natural language processing: related techniques can be found in theorem proving, knowledge representation research and theory of data types.

訳文

［人間］ 歴史的に見ると，これらの文法形式は，計算言語学，形式言語，自然言語処理の各分野でそれぞれ個別の研究がなされてきた結果として生まれたものである（このようなアプローチは，定理証明，知識表現の研究，データ型の理論などの分野でも似たようなアプローチが見られる）．

［機械イ］ 歴史的に，これらの文法形式主義は計算言語学と形式言語そして自然言語処理の別々の研究の成果である；定理証明と知識の表現の研究とデータ型の理論に関係がある技法を見出すことができる．

［機械ロ］ 歴史的に，これらの文法形式主義は計算言語学，形式的な言語学および自然言語処理の別の研究の成果である；関連の技術はデータタイプの定理の証明，知識の表示の研究，および理論に見つけられ得る．

［機械ハ］ 歴史上，それらの文法形式は計算言語学，形式言語学および自然言語処理における個別の研究の成果である；関連する技術は定理証明，知識表現研究およびデータタイプの理論に見い出すことができる．

（a） 英語--〉日本語

原文

これらの新しい計算言語学理論に共通していることは，文法機能が作り上げる局所構造の組み合わせで全体構造を再帰的に定義する点である．

訳文

［人間］ Common to all of these new computational linguistics theories is the fact that the overall structure is defined by rccursivcly and iteratively combining the local structure in terms of grammatical functions.

［機械A］ Common to the computer linguistic theory with new these is points to recurrently define whole structure by combining the limited part structure a grammatical function makes up.

［機械B］ A point that whole structure is defined recursively in the combination of the local structure that grammatical structure makes is common to these new computational linguistics theories.

［機械C］ The point that is common to these new computational linguistics theory is defining whole structure in a combination of local area structure that grammar function makes up reflexively.

（b） 日本語--〉英語

注 1. 人間の翻訳者は，翻訳過程で内容についてのコメントが与えられた．
2. 機械システムは，辞書のチューニングなどが許された．

図-3 機械翻訳の例

軟性に富んだシステムの実現が望まれている．

1.1.3 翻訳の実例

次に，機械翻訳 (Machine Translation) の実例を示そう．もちろん“機械”は計算機である．**図-3**は，わが国で開発されたいくつかの実用化レベルのシステムで実験した例を示している[1]．（a）は英語から日本語へ，（b）は日本語から英語への翻訳である．それぞれにおいて，参考までに与えられた原文をまず人が翻訳し，ついで異なったMT システムの翻訳した結果が示されている．機械システムには，少しばかり事前のチューニングが許されたようだが，（a）については，機械の健闘ぶりが窺えよう．人に比べて，大幅な不具合は見当たらない．平均してこの程度の翻訳ができれば，機器の取扱い説明書や商品のカタログなど，分野によっては大いに実用化が期待されよう．一方（b）の結果を見ると，“共通していることは，～点である”に関して，システムごとに訳文の構造が大きく異なっている．この文の英訳は，筆者にとっても容易でないが，処理結果を見ると原文の大意はほぼ翻訳されている．

以上，現在の MT 技術の一部を紹介した．やはり当面の対象は，この例のような科学技術文献あるいは取扱い説明書やカタログのような分野であろう．文学作品のような翻訳自体に芸術性さえ求められる分野は，次節で述べる AI 技術が格段に進歩した，かなり将来の目標であろう．

1.2 自然言語処理の歴史

NLP の歴史は，二つの研究開発の流れ，

＊ MT など実用化を目指す研究，
＊ 人の言語理解過程さらには知能の解明を目指す AI 研究

に，それらを支える

＊ 言語理論

が絡み合って，複雑な様相を呈している．これらの流れを，当時の社会的背景にも目を向けながらたどってみよう[2)~4)]．

1.2.1 MT 志向の流れ

世界で最初の計算機は，1945 年にペンシルバニア大学で開発された **ENIAC** とされている．ところがその翌年，イギリスのバークベックカレッジの A.D. ブースとロックフェラー財団の W.

ウィーバーは，いち早く計算機による電子辞書の構想や翻訳処理の可能性を議論した．驚くべき先駆性といわざるをえない．

1950年代後半になると，次々にMTの研究グループが発足する．アメリカでは，NSF (National Science Foundation) の助成を得て，MIT，ジョージタウン大学，RANDコーポレーションなどが研究を始めた．これに一層拍車をかけたのが，1957年，旧ソ連による人工衛星スプートニクの打ち上げである．それまで科学技術における優位性を信じていたアメリカにとって，大きな驚きだった．そこで，ソ連の大量の科学技術文献の翻訳が重要な課題となった．

ヨーロッパでは，上に述べたブースのグループやフランスのグルノーブル大学CETA (現在のGETA) などがあげられる．わが国でも，1958年に九州大学が，またほぼ同じころ通産省電気試験所 (現在の電子技術総合研究所) が開発を始めた．この頃は，ハードウェアやソフトウェアが未熟であったにもかかわらず，研究者たちが果敢にMTシステムの実現に取り組んだ時代といえる．

順調に成長を続けるかにみえたMT研究は，1966年，大変衝撃的な報告に出あった．NSFの組織した，いわゆる**ALPAC** (Automatic Language Processing Advisary Committee) のレポートである．これは，MTの研究状況と周囲の状況から，(当時を起点にして) 近い将来MTが有用であると予測できる展望がない，つまり必要性，経済性その他の理由から人の行う翻訳に比べて利点がない，というものである．このレポートの影響は大きく，特にアメリカではこれを機にMTの研究は急速に衰退してしまい，ヨーロッパやわが国にもこの傾向が波及した．

このような逆境にもかかわらず，MTの必要性の高い国では地道に研究が続けられた．カナダでは，公用語として英語とフランス語が用いられるため，政府はその翻訳作業に悩んでいた．モントリオール大学では，1962年以来，英仏翻訳TAUMプロジェクトを進めていたが，1976年には，ついにTAUMMETEOという天気予報のシステムの実用化に入った．またヨーロッパ共同体ECの事務部門では，加盟国の間の文書の翻訳作業におよそ2千人もの翻訳者が携わっていると聞くが，その労力と予算は大変なものである．1976年にはジョージタウン大学のシステムを改良した，SYSTRANとよばれるシステムを導入し，MTの実用化に乗り出した．

このころわが国でも，文書の機械処理に関して大きな出来事が起こった．1979年のワープロの出現である．わが国では，**1.1.1** で触れたように，かな文字に加えて多くの漢字を用いるため文書の機械処理が遅れていた．しかし，かな文字の文を漢字とかなの混合文に変換する技術が進歩し，ついにワープロが市販された．ご承知のようにワープロは急速に普及し，“文書の機械処理”という概念を広く社会に浸透させる大切な役割を果たした．

1970年代末期から1980年代初頭にかけて，再びMTシステム開発の機運が，今度はヨーロッパと日本で盛り上がってきた．ヨーロッパでは，上述のECが1982年からEUROTRAプロジェクトを開始し，英語，フランス語，ドイツ語，イタリア語，オランダ語，デンマーク語，およびギリシャ語の，7カ国語間の相互翻訳を目指した．またわが国では，科学技術文献に関する大量のデータベースを輸入しながら，翻訳作業がネックとなって輸出のほうがはかどらないという国際問題を抱えていた．これに，ワープロで力をつけた民間企業が次のレベルのNLP技術を開発したいという機運も重なった．そこで科学技術庁のMuプロジェクトが発足し，英日と日英のシステムの実現を目指した．

1980年代の処理技術は，1960年代に比べて，単語辞書を格納したり，それを検索したり，あるいは構文を解析したりする点で，長足の進歩を遂げたといえよう．これには，言語理論やプログラミング技法の進歩も大きく貢献している．しかし翻訳の質の点では，着実な進歩を遂げているものの，人の翻訳者と比較した場合，やはりまだ道程を感じざるをえない．大きな課題の一つは，**1.1.2** でも述べたように，言語の性質に関する調査，そしてそれに基づく言語データの蓄積である．数十万におよぶ単語の文法データ，より詳細な構文規則，それに各語の意味もしくは概念とそれらの間の関係のデータである．

わが国では，1985年に電子化辞書研究所が官民の協力で設立され，MTを含む各種のNLPの言語データの蓄積に努めている．NLPの質の向

上には，このような地道な積み上げこそが，遠回りのようにみえて実は一番の近道と筆者らは考える．

1.2.2 AI 志向の流れ

実用化を意識する MT 志向の場合は，広い範囲をカバーできる，大規模な NLP システムを目標とする．品質の高い翻訳には言語の理解が必要であるが，理解の問題よりも大規模化に重点が置かれる．それに対し計算機が言葉を理解することに興味をもつ AI 志向の場合は，広い範囲にわたっての体系化は難しく，ある分野で意味や知識を深く追求しようとする．そこで意味や知識を計算機内部でどのように表現し処理するかに注目して，歴史の節目をとらえてみよう．

1956年，ダートマス大学で“知的な動作をする計算機”が議論され，ここで初めて“人工知能”という名称が与えられた．1961 年には，MIT の M. ミンスキーが“人工知能に向けての歩み”という論文を発表し，これにより AI 研究の枠組みが明確になった[5]．

1960 年代末，M. R. キリアンによって**意味ネットワーク** (semantic network) の手法が提案された[6]．たとえば“指は手の部分である”という知識は，**図-4** に示しているように，“手”と“指”という二つの節（ノード）を作って，それらの間を PART-OF というポインタで結びつける．このようにして複雑な意味もしくは概念の間の関係をネットワークでとらえようとするものである．

1970 年代の初めに，自然言語の意味理解の研究に大きな衝撃波が走った．T. ウィノグラードの**SHRDLU** とよばれるシステムの出現である[7]．グラフィック装置に表示されている積み木に対して人が言葉で指示すると，システムが指示に従って積み木を動かす．システムは人と会話することもでき，言語と行動の統合的なシステムといえる．SHRDLU は，ALPAC レポートによって暗い雰囲気になっていた NLP の研究者に，光明を与えてくれた．そして現実世界との対応のもとで意味情報の取扱いに新しい展望を与えてくれた．

1970 年代は，知識表現や意味処理技術が大きく進展した時代といえる．R. C. シャンクは**概念依存法** (conceptual dependency) とよばれる知識表現を提案した[8]．これは文の意味を支配する動詞に注目し，それを修飾する名詞などを図式的に結びつけようとするものである．たとえば，“John hits his little dog yesterday” という英文において，John, dog, yesterday はそれぞれの役割をもって hit を修飾しており，その役割を図式的に示すものである．シャンクはこれを出発点として，さらに大きな知識の塊をとらえる形式や，与えられた目標を解決するためのプランニングなどについて研究を進め，自然言語理解のみならず，心理学の分野にも大きな影響を与えた．

わが国でもシャンクに先駆けて，九州大学の栗原，吉田が概念依存と意味ネットワークを総合した形で研究を進めていた[9]．しかし多くの論文が日本語で発表されたため，国際的に日の目を見るに至らなかったのは残念なことである．

このころの AI 研究は，いろいろな人がおのおののアイデアに基づいてできるかぎり人に近い振舞いをするシステムの構築を試みた，意気盛んな時期といえよう．しかしながらその多くは個別的で，特定の問題にしか応用の期待できないものであった．ミンスキーはそれらを概観し，統一的な知識表現を目指して，1975 年に**フレーム理論** (frame theory) を提案した[10]．フレームはその名のとおり枠組みのことで，個々のデータは“スロット”とよばれる部分に格納される．スロットには通常こうであろうと予測されるデータ，すなわち“デフォールト値”が入る．たとえば“椅子”というフレームの“用途”というスロットには，“人が座る”というデフォールト値が入る．一つの小さなフレームは，それ全体がより大きなフレームのスロットに入り込むことも可能で，小さな知識が次第に大きくまとめられる．**図-5** に具体例を示している．“部屋”というフレームが，小さなフレームからまとまっていく様子が示されている．

フレーム理論は，その後の AI 研究に大きな影響を与えた．言語理解の分野においても，しばしば知識表現にフレームが用いられる．なお次回の講座で紹介する“格フレーム”とよばれる考え方は，知識表現としては，フレーム理論を言語部門へ応用したもの，とみることもできよう．

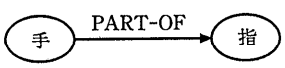
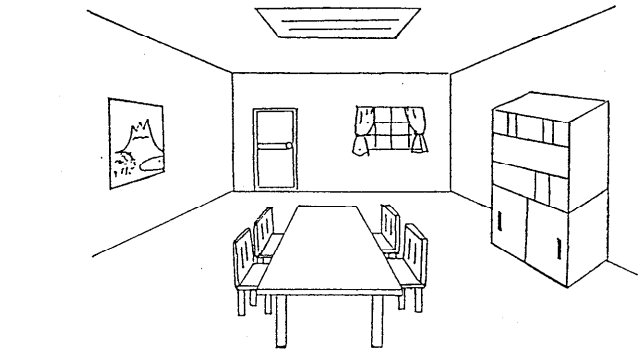
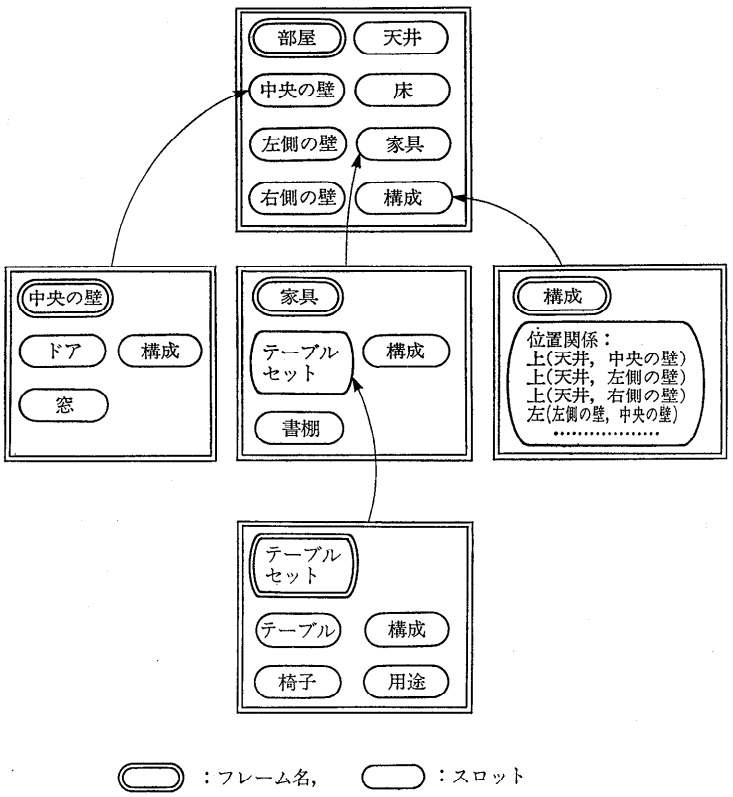


図-4 意味ネットワーク



(a) 概観図



(b) 部屋フレーム

図-5 フレーム構造・・・部屋

1.2.3 言語理論

言語学の歴史は古く，対象とする分野も音韻論，形態論，構文論，意味論，語用論など，広い範囲にわたっている．構文論に関して，特に機械処理に結びつきの深い理論を一つ取り上げて，その流れを眺めてみよう．

1957年にN.チョムスキーの提案した**生成文法** (generative grammar) は，言語学に革命をもたらしたといわれている[11]．この文法は，機械処理に大変都合のよい考え方で，基本的な文は X→Y（XとYは一つの記号とは限らない）という形式の“書換え規則”の集まりでとらえる．たとえば，文Sが名詞句 NP と動詞句 VP から構成されていることを，S→NP VP のような書換え規則で表す．また複雑な文は，“変形”とよばれる操作によって基本文から導くことにする．たとえば英語の受動文は能動文から“受け身変形”によって導かれる，と考えた．特に書換え規則の集まりでとらえる部分は**句構造文法** (phrase structure grammar) とよばれ，数理的観点からも詳しい研究がなされた．生成文法は，言語学の分野ではもちろんのこと，NLP の分野でも高い評価を受け，世界的に広まった．

生成文法の歴史は，変遷の歴史といっても過言ではない．1965年にチョムスキーは，初期の理論を改めて**標準理論**を提唱した[12]．**図-6** に示しているように，文の構造を表面上に現れた表層とその背景にある深層の二つのレベルに分け，句構造文法で“深層構造”を生成し，それに変形を施して“表層構造”を導くものである．その後さらに改良を加えて，現在は**統率・束縛理論**とよばれるものへと進展している．これらを通じてチョムスキーと彼の学派の目指しているのは，英語や日本語に固有の文法ではなく，人類に共通な**普遍文法** (universal grammar) の確立，ということのようである．人は生まれながらにして言語の機構を備えており，それはいくつかの原理の体系によって説明される．日本語や英語は，それらの原理のパラメータの値によって定まる，と考えるものである．

以上，NLP の歴史を眺めてきた．1990年代は，

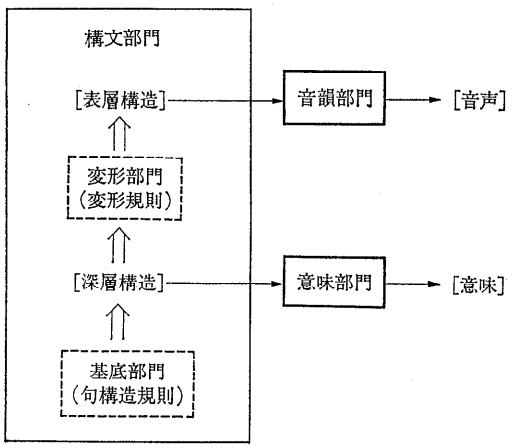


図-6 標準理論の構成

これまでに培った構文情報中心の NLP 技術に AI 研究の成果としての意味処理技術が加わり，さらに計算機処理を意識した言語理論がそれらを支援して，特定の分野に関して質の高い実用化システムが期待される．

1.3 本講座の構成

本講座の構成を述べておこう．

第2章は言語理論で，機械処理の立場から構文論と意味論を解説する．構文論はチョムスキーの標準理論を述べる．特に“S+V+O”のような学校文法になじんだ読者に，“X→Y”型の数理的な概念を導入する部分を詳しく説明する．意味論は，語の意味を概念とみなし，まず単位としての“意味素性”を導入する．そして文の意味を語の概念とそれらの間の結びつきとみなし，この意味構造を格フレームでとらえる考え方を解説する．

第3，4章は言語解析で，与えられた文を形態素，構文，意味および文脈の各レベルで解析する手法を示す．紙面の都合により，第3章で，“有限オートマトン”とよばれる手法を用いる英語の形態素解析と，かな漢字変換を中心にして日本語の形態素解析を述べる．第4章では，まず言語処理とプログラミング技法の歴史について少し触れる．続いて構文解析として，“トップダウン/ボトムアップ”と“シリアル/パラレル”という基本概念を示す．そしてトップダウン・シリアル型の一つの解析法に注目し，やや詳しく計算機内部でのプログラムの振舞いを観察する．さらに，**確定節文法** (definite clause grammar) とよばれる表現形式で，Prolog とよばれるプログラミング言語の上に簡単な解析システムを実現する．意味解析は，この構文解析システムと新たに導入する**単一化文法** (unification grammar) に，意味素性や格フレームを導入して行う．最後に，文脈解析の考え方を例示する．

第5章は言語生成で，ある意味内容を深層構造から表層構造を経て文として生成する手法を示す．深層構造については，非言語的な思考内容から深層の格フレームを抽出する，筆者らのアイデアを例示する．表層構造については，深層格フレームの列から文章を生成する手法を述べる．

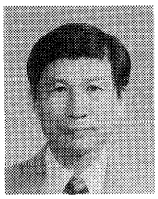
第6章は言語データで，機械処理に必要な文法，辞書，および意味データについて解説する．文法と辞書データについては，中学程度の英文テキストを素材にして構文規則と語彙データを実際に作成する手法を述べる．さらに，得られた構文規則と語彙データに対して意味データを作成する．

それでは次回は，機械処理に適した文法論と意味論について解説しよう．

文　献

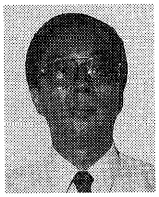
1) 機械翻訳システム調査専門委員会 (編)：機械翻訳の開発と実状に関する実態調査，日本電子工業振興協会 (1989).
2) 和田　弘：機械翻訳の現状と将来，情報処理，Vol. 6, No. 3, pp. 126-128 (Mar. 1965).
3) 田町常夫：機械翻訳の概要と歴史，情報処理，Vol. 26, No. 10, pp. 1140-1147 (Oct. 1985).
4) 長尾　真：計算言語学の歴史と展望，情報処理，Vol. 27, No. 8, pp. 855-861 (Aug. 1986).
5) Minsky, M.: Steps toward Artificial Intelligence, Proc. IRE, Vol. 49, pp. 8-30 (1961).
6) Quillian, R. A.: Semantic Memory, in Minsky, M. (ed.): Semantic Information Processing, MIT Press (1968).
7) Winograd, T.: Understanding Natural Language, Academic Press (1972).
8) Schank, R. C.: Conceptual Dependency: A Formalism of Natural Language Understanding, Cognitive Psychology, Vol. 3, No. 4, pp. 552-631 (1972).
9) 栗原俊彦：自然言語の機械処理，情報処理，Vol. 14, No. 4, pp. 267-274 (Apr. 1973).
10) Minsky, M.: A Framework Representing Knowledge, in Winston, P. H. (ed.); The Psychology of Computer Vision, McGraw-Hill (1975).
11) Chomsky, N.: Syntactic Structures, Mouton, The Hague (1957).
12) Chomsky, N.: Aspects of the Theory of Syntax, MIT Press (1965).

(平成5年6月4日受付)



岡田 直之（正会員）

1964年東海大学工学部卒業．1966年九州大学大学院工学研究科修士課程修了．同年同工学部助手，1976年大分大学工学部助教授，1978年同教授を経て，現在九州工業大学情報工学部教授．工学博士．人工知能の研究に従事．電子情報通信学会，人工知能学会各会員．



中村 順一（正会員）

1979年京都大学工学部卒業．1982年同大学院工学研究科博士後期課程中退．同年京都大学工学部助手，1989年九州工業大学情報工学部助教授．工学博士．自然言語処理，音楽情報処理の研究，計算機ネットワークの管理に従事．電子情報通信学会，ソフトウェア科学会，日本認知科学会，Association of Computational Linguistics 各会員．