

自然言語に内在する仮説生成機構

安藤 司文

長崎大学工学部機械システム工学科

概要

本研究では、人間の頭の中にある概念構造を記号化して取り出したものを意味言語となすけ、この意味言語を用いて、自然言語の中に存在する知的なメカニズムを解明しようとしている。

本論文では、次のことが明らかになった。即ち、ある自然文を意味言語に変換し、その意味言語において下位-左側優先のルールに従って、格構造を論理構造（IF論理構造）に変換し、さらに固有名詞を変数名詞に、また時間、空間に関する単語をデフォルト値に変更することによって、この自然文からいくつかの仮説が生成できる。

The mechanism for forming hypotheses in a natural language

Shimon Ando

Dept. of Mechanical Systems Engineering,
Faculty of Eng., Nagasaki Univ.

Bunkyo-machi, Nagasaki 852, Japan

Abstract

A meaning language (MI-language) has been proposed, where conceptual structure in the human brain is symbolized and taken out. With utilization of the meaning language, the intellectual mechanisms embedded in a natural language has been studied.

In this paper, it is found that hypotheses formed from a natural sentence are generated by changing case-structures of the meaning language to logical structures (IF-logical structures) in accordance with the priority rule (low-right-rule) and by altering proper-nouns to variable-nouns and the value of words concerning about space and time to default-value.

1. はじめに

人間は自然言語を用いて、質問応答、推論、知識獲得、学習、翻訳、物語理解などの知的活動を行っている。そのため、自然言語の中にこれらの知的活動を支えているメカニズムが存在すると考えられる。しかしこれまでの自然言語の研究は文法理論、知識工学、機械翻訳、エキスパートシステムなどでの各分野に限定して研究が行われてきた(1)。本研究では、上述の分野について個別に研究するのではなく、総合的に研究を行うことを提案している。

このことは問題を複雑にするのではなく、むしろこれまでの自然言語の問題を単純に解き明かすことができることが分かってきた。ただし、本研究では、自然言語をそのまま用いるのではなく、人間の頭の中に存在する概念構造を記号化して取り出したものを意味言語と名付けて、自然言語の代わりに用いている。

自然言語は個別言語の文法や話し手の運用上の判断や状況認識によって表現(文字列)が著しく影響を受けるので、自然言語をその倣用いることはできない。意味言語はコンピュータ上で自然文の意味が正確に表現でき、しかも上述の各種の知的処理が簡単に行える。Schankのように、自然言語の他に、別の形式(概念依存構造)で意味を記述し直す(2, 3)必要はない。これからの一連の論文でその結果について報告する。

推論は知識に基づいて行われるのであるから、推論を行うためには、知識がなければならない。しかし、知識を獲得するためには、まず仮説を生成しなければならない。人間は生活している環境から、種々の情報を獲得するが、その情報から種々の仮説を生成し、その仮説を実世界の中で検証し、その検証に耐えて、その人間にとって有益であると判断すると、その仮説を知識に格上げして蓄える。人間は日常の生活の中で絶えずこの知識を用いて推論して、推論による仮想的な世界を頭の中に描き、現実の世界と比較検討しながら、知識の信頼性を確認している。信頼性が低い場合はより信頼性の高い知識に更新するよう絶えず努力している。エキスパートシステムにおいて、システム自身がこのように自動的に仮説を生成し、その

仮説を検証しながら、信頼性の高い知識を獲得することができれば、より信頼性の高いシステムを構築することができる。本研究の目的はコンピュータで制御されるシステムが人間と同じ言語を持ち、人間とシステムが自然言語で情報を交換しながら、人間と同じプロセスでシステムの知能を向上させて、信頼性の高いシステムを構築することである。

意味言語の普遍文法、自然言語から意味言語への変換、即ち、自然言語の意味解析、や意味言語から自然言語の生成、パターンマッチングによる質問応答、機械翻訳などのについては別報で詳しく述べるとして、本論文では、自然言語から変換された意味言語を用いて、仮説が生成されるプロセスについて説明する。

2. 意味言語の提案

本研究では、前述のように自然言語は話し手の運用上の判断や状況認識によって、複雑に影響を受けるので、そのようなものに影響を受けないもの、即ち、話し手の頭の中にある概念構造、を記号化して取り出し、それをメタ言語として用いている。話し手が自然言語を表出するには、まず話し手の頭の中には表現したいと考えている概念構造(意味言語)が、図1に示すように、存在しなければならない。本研究では、自然言語はこの頭の中にある概念構造が話し手によって加工され、個別言語の文字列に変換されたものであると考えている。聞き手に話し手の概念構造と同じ概念構造が取り出されたときに情報伝達が完結する。情報の伝達は話し手の意味言語と聞き手の意味言語との関係によって決まり、自然言語には基本的に関係がない。自然言語は聞き手に話し手の意味言語と同じ意味言語を取り出させるための道具にすぎない。

ここで意味言語から見た自然言語の三つの原理を挙げる。

原理X: 自然言語の言語構造そのものには一切の曖昧性はない。

原理Y: 自然文で表現される単語および助詞は意味を確定すると同時に他の多くの概念構造と識別するためのマークである。

原理Z: 必要性のない単語及び助詞はできるだけ

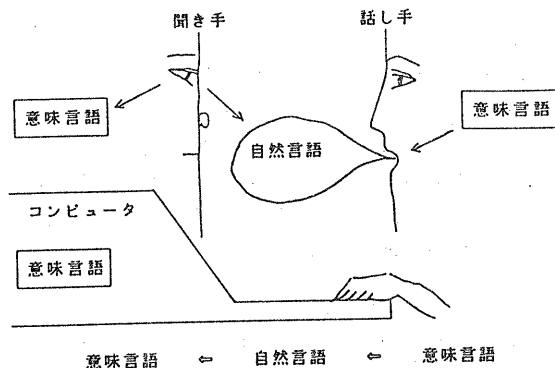


図1 意味言語と自然言語との関係

表現を省略し、表現されている単語及び助詞はそれぞれ重要な意味を持つものとして評価される。

この自然言語の原理については、折りにふれて説明する。

3. 意味言語の基本的な考え方

- (1) 概念構造を図式的に分かり易くすること。
 - (2) 意味言語から自然言語が生成されるプロセスが容易に分かるようにすること。
 - (3) 話し手が状況認識や運用上の判断によってどのように意味言語を加工したかが分かるようにすること。
 - (4) コンピュータ上のデータ構造と概念構造との対応付けを分かり易くすること。
- などから、次のような要素と記号を用いて意味言語を表現する。意味言語の構成要素としてMWとPSを設定した。

3.1 要素MW

本意味言語の文法では、概念を表す単語とその単語が概念構造の中でどのような役割、位置を持っているかを示すマークとしての助詞しかない。動詞、形容詞は後述の意味フレームを代表する単語で、意味解析のときに用いる。主語、目的語、補語などは概念構造のある特定な位置の単語に付けられた役割で、本文法ではほとんど何の役割も持っていない。単語以外は全て助詞で、その為に多くの助詞を設定している。本論文ではここで必要なものだけをその都度説明する。

MW (META WORDの略)には概念を表す単語が格納される。後で述べる構造文では分かり易くするために単語を日本語で表現しているが、実際にはコンピュータの中で、後で述べるデータ文(表1)に示すように、世界共通に使用できる意味コードを用いている。単語に付随する各種の助詞もこの要素に格納される。要素MWを()の記号で表す。他の要素MWや要素PSと結合するために結合手を持っているが、これを⇒や⇓で表す。要素MWは図2に示すように、上下2本、左右2本の合計4本の結合手を持っている。その他、比喩、本音などを表現するための文や副詞句を接続するために専用の結合手もっているが、本論文ではこの点に触れないので、省略する。この結合関係はコンピュータ上では、互いに相手の要素の番号(アドレス)を持つことで表現される。“あの”、“その”、“この”などの冠詞も助詞として取り扱っているが、冠詞は半角で()の左側に、格助詞は右側に示す。従って、
”あ(太郎)が
”のように表記する。

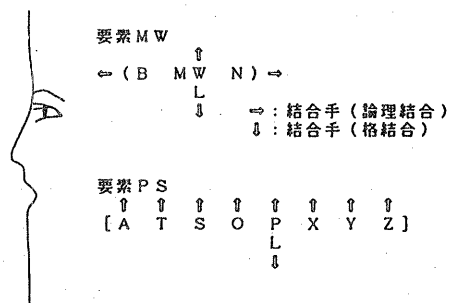


図2 意味言語を構成する要素、MWとPS

3.2 要素PS

単語がいくつか集まって文章ができるが、基本的な意味単位を構成する文章である基本文として、要素PS (PRIMITIVE SENTENCEの略)を設定した。この要素PSは次の五つの格を持っている。

- A格: 主格 Agent case
- T格: 時間格 Time case
- S格: 空間格 Space case
- O格: 対象格 Object case
- P格: 述語格 Predicate case

さらに、次の三つの補助的な格、即ち

X格：補助格 Auxiliary case
 Y格：応答格 Yes-No case
 Z格：全体格 Zentai case
 A格は（誰），（who）などを格納するMWと結合する格，T格は（いつ），（when）などの時間に関する単語を格納するMWと結合する格，S格は（どこ），（where）などの空間に関する単語を格納するMWと結合する格，O格は（何か），（what）などを格納するMWと結合する格で，P格は（どうする），（how）などの述語（動詞）を格納するMWと結合する格である。つまり基本本文PSは，who, when, where, what, howを示す格から構成されている。何故（why）は別報で詳しく述べるが，PSとPSとの論理関係，つまり，（何々）IF ⇒ ならば（何々）

という原因，理由などを表すもので，これもPSとの関連で考えると，1H5Wが意味の基本単位である。

X格，Y格，Z格はいずれも補助的に用いられるもので，X格は，英語で it~that~ という構文で that 以下の文を保管したり，複雑な語順の文を生成するときや，構文解析や意味解析で，そのときには判断ができないときに一時保管するとき用いるものであるが，本論文ではこの点については触れない。Y格は“はい”，“いいえ”，“Yes”，“No”などの応答のときに使う単語を格納するMWと結合する格である。文章全体を一つの単語として取扱い，それをMWに埋め込むことがあるが，Z格はその埋め込み文の性格を表現する単語を埋め込むときに用いる格である。これについては後で少し述べる。

要素PSはA，T，S，O，P，X，Y，Zの8個の格と結合する要素MWと下位にある要素MWと結合するために合計9本の結合手をもっている。即ち，人間の頭の中にある概念構造は図2に示すように，上下左右の4本の結合手を持っている要素MWと9本の結合手をもっている要素PSによって構築される。

図2において，B（BEFOREの略），N（NEXTの略），L（LOWの略）及びMWは要素MWのメンバで，結合相手の番号（アドレス）を格納する。互いに相手の番号を格納することによって，結合関係が保証される。MWには結合する上位の要素の番号が格納される。

人間の頭の中に構築されている概念構造を図式的に示したものが，図3の（a）である。これを構造文と呼ぶことにする。この概念構造はコンピュータの中では（b）のようなデータ構造で表現される。これをデータ文と呼ぶことにする。本論文の内容と関係ないものについては省略し，結合関係だけを示した。データ構造（データ文）は要素PSとMWに別れている。

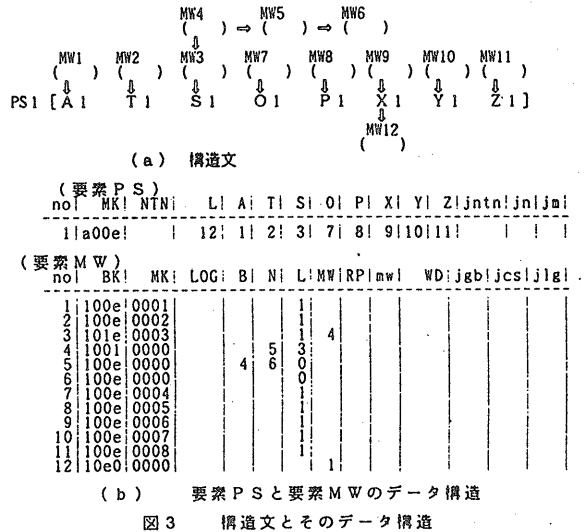


図3 構造文とそのデータ構造

PS1のA格はMW1を結合しているので，PS1のメンバAにMWの番号1が書き込まれており，MW1は下にあるPS1のA格に結合されているので，MW1のメンバLにそのPSの番号である“1”が書き込まれている。MW5は前にあるMW4と後ろにあるMW6と結合されているので，MW5のメンバBに“4”が，メンバNに“6”が書き込まれている。MW3は上のMW4と結合しているので，MW3のメンバMWに“4”を，MW4は下のMW3と格結合しているので，MW4のメンバLに“3”が書き込まれている。MW3は下にあるPS1と結合しているので，メンバLに“1”を指定するが，数値だけでは要素の種類が判別できないので，メンバBKの16進数の4桁の下2桁を使って，要素の識別を行う。“1”がMWで，“e”がPSとして，1桁目が下に結合されている要素を，2桁目が上に結合されている要素を示す。MW3は上のMW4と下のPS1に結合されているので，BKが“001e”となっている。このような表現で（a）で示した

結合関係が規定でき、どの要素からでも、上下左右のどちらの方向にも自由に要素をたどることができる。MWのメンバWD (WORDの略)には単語の意味コードが格納される。

要素MWと要素PSが縦方向に結合(格結合)したり、横方向に結合(論理結合)して、いろいろな意味を持つ概念構造が構築される。我々人間が普段使用している意味には、それに応じた概念構造がある。あるひとまとまりの概念構造を意味フレームと呼ぶが、動詞、形容詞はそれぞれに対応した意味フレームを持っている。

意味フレームは概念辞書に登録されている。本意味言語では自然文が表現する意味を動詞、形容詞などの意味フレームを用いて文章全体の意味の枠組を作り、そのスロットに単語を埋め込んで意味を確定している。この点を説明するためには、意味言語の言語構造について説明しなければならないが、詳しくは別報で述べるとして、ここでは本論文で必要なものだけ簡単に説明する。

4. 意味言語の言語構造

4.1 格構造

“太郎”、“今日”、“グラウンド”、“ボール”、“投げた”などの単語はそれぞれ特定の概念(意味)に割り当てられた記号である。それが次に示すように“太郎”が要素PS1のA格に、“今日”がT格に、“グラウンド”がS格に、“ボール”がO格に、“投げた”がP格に格結合すると

{太郎が今日グラウンドでボールを投げた}
 という自然文が表現する新しい意味(概念構造)が生成される。これを構造文で示したのが、図4である。“投げた”の“た”は動作の完了を表す動詞の活用語尾であるが、本意味言語では格助詞として取り扱っている(詳しくは別報で述べる)。また述語の“投げた”は意味フレームを表す単語であるから、“太郎”などの単語と明確に区別すべきであるが、意味言語の基本的な言語構造を分かり易く説明するために、こゝしばらくは単なる単語をして取り扱うことにする。

単語が埋め込まれた格はその文章から取り出されて、次の文章に埋め込まれることによって、概念構造が多層化して、より複雑な意味が表現でき

る。格の取り出しは、図4で示すようにPS1の下に要素MW6を設け、それに取り出した格の内容を埋め込むことであるとする。このようにしたのはある文からある格を取り出すことは、取り出された内容をその文が規定している概念で新しく規定し直すことであると考えたからである。

例えば、{太郎が今日グラウンドでボールを投げた}という文から“太郎”を取り出すと、{今日グラウンドでボールを投げた太郎}になるが、これを構造文で示すと、図4の(a)のようになる。MW1の(太郎)は単なる“太郎”であるが、MW6の“太郎”は<今日グラウンドでボールを投げた太郎>というように、この文章で規定され直された“太郎”であると考ええる。

この構造文から自然文は、↓を取り除き、図示した順番に文字列を並べると、

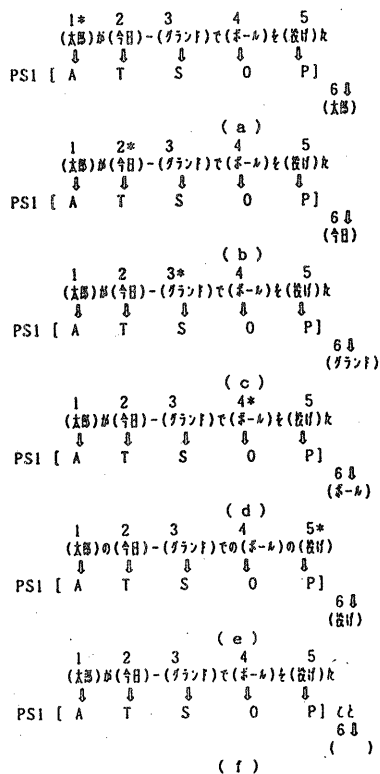


図4 格取り出しのプロセス

[(太郎)が(今日)(グラウンド)で(ボール)を(投げ)た] (太郎)

これを半角を全角に変換して、()や[]を取り除くと、

{太郎は今日グラウンドでボールを投げた太郎}

になる。

ここで、“太郎”が二回でくることが、必要性のない単語は表現しないという自然言語の原理Zによって、重要性の低い方の“太郎”を表現禁止にする

下位になるほど重要性が高いので、上位の“MW1にある“太郎”を表現禁止にして、次のような自然文が得られる。

{今日グラウンドでボールを投げた太郎}

構造文では表現禁止のマークとして、*印を付けた。この表現禁止というのが本意味言語の基本的な考え方である。Chomskyの統率、束縛理論(Government-Binding Theory)(3~5)では痕跡(trace)、あるいは空範疇としているが、この意味言語ではただ単に単語の表現が禁止されているだけであると考えている。単語が埋め込まれた格の内容は全て取り出すことができる。図4にはA、T、S、O、Pなどの格および文全体を取り出したときの様子を構造文で示した。この構造文から自然文を生成すると、次のようになる。表現禁止の単語を で示す。

A格の取りだし：

{太郎が今日グラウンドでボールを投げた太郎}

T格の取りだし：

{太郎が今日グラウンドでボールを投げた今日}

S格の取りだし：

{太郎が今日グラウンドでボールを投げたグラウンド}

O格の取りだし：

{太郎が今日グラウンドでボールを投げたボール}

P格の取りだし：

{太郎の今日グラウンドでのボールの投げた投げ}

全体格の取りだし：

{太郎が今日グラウンドでボールを投げたこと}

A、T、S、O格はこのように容易に取り出すことができるが、図4の(f)に示すように文章全体を取り出して、それをあたかも一つの単語のように取り扱おうとすることがある。文章全体を取り出すのであるから、これを全体格の取り出しとせずけた。MW6の()には、何も書き込まれていないが、上にある概念構造全体を受けていると考える。{太郎が今日グラウンドでボールを投げたという事実}や{太郎が今日グラウンドでボールを投げたというわけ}などのように“~という事実”や“~というわけ”などのように文章全体の状

態を示す単語はZ格に埋め込まれるが、そのZ格を取り出すのではなく、文章全体を取り出すと考える。これを示すために[]の右側に“こと”という助詞(全体助詞)を示した。

述語(動詞)はある意味フレーム全体を代表する単語であるから、これを取り出すことが、文章全体を取り出すこととはほぼ同じ意味になる。この場合、格助詞は“が”は“の”に、“で”は“で”の”に変化したり、述語格の格助詞は除去され、過去、丁寧、否定などは表現できなくなる。

単語が埋め込まれていない格も取り出すことができる。{投げたひと}、{投げたとき}、{投げたところ}、{投げたもの}、{投げ}、{投げたこと}などの文章を前述の説明と対応して考えると、{投げたひと}の”ひと”は{投げた}のA格から、“とき”はT格から、“ところ”はS格から、“もの”はO格から取り出された単語であることが分かる。従って、“ひと”、“とき”、“ところ”、“もの”は{投げた}という文では表現されていなくても、実際には

{ひとがときところでものを投げた}

という意味に規定されていることが分かる。これでは日本語としては少し不自然であるので、“ひと”の代わりに“誰か”、“とき”の代わりに“いつか”、“ところ”の代わりに“どこか”、“もの”の代わりに“何か”を用いると、{投げた}という文章は

{誰かがいつかどこかで何かを投げた}

という意味であることが分かる。つまり、これらの単語は や*で示すように表現禁止になっているだけで、デフォルト値として、上述の単語が埋め込まれていることが分かる。これらの単語は人間にとってはいちいち言わなくてもわかるので、自然言語の原理Zによって、表現が禁止になっていると考える。

4.2 論理構造

要素MWは一つの概念を表す。要素MWが要素PSの格と結合すると格構造を構成する。この関係を↓印で表したので、別に、縦構造と呼ぶことがある。要素MWが⇒の印で横に結合すると、これを横構造と呼ぶことにする。別に論理構造と呼ぶことがあるが、これは記号論理学の“and”

、" or"、" not"、" ならば"などの結合関係に比較的思考方が近いので、このような表現を用いるが、形式論理とはほとんど関係がない。誤解が生ずるようであれば、横結合、横構造という表現に統一してもよいが、論理関係は一般的にルーズに用いられているので、ここでも論理構造、論理結合という表現を用いる。論理関係は本意味言語では、次のように表記される。

() IF ⇒ ならば ()

⇒の左側に 論理関係の種類"IF"を、右側に論理助詞"ならば"を半角で示す。本意味言語では論理構造の種類を、含意とそれ以外に大きく分けている。それ以外には、" AND"、" OR"、" NOT"、" THAN"があり、" 複数"、" ~の"、" ~さえ"、" ~すら"、" ~でも"、" ~だけ"、" ~のみ" " ~しか"は一種のAND関係であるとしている。

含意関係としては、THEN (~であり、~である)、IF (~ならば、~である)、BUT (~であるが、~である)、EVER (~であるとしても、~である)、AS (~であるので、~である)、FOR (~のために、~である)、BY (~によって、~である)の7種類を設定している。詳しくは別報で述べる予定である。

⇒の両脇に表記する論理関係の種類の記事号や論理助詞は説明に必要でない場合は省略している。

概念を表現する要素MWが、基本文である要素PSを核として、縦、横に多層に結合され、複雑な概念構造が構築される。つまり、概念構造は多層格-論理構造である。我々人間が普段使用している意味には、それに応じた概念構造がある。あるまとまりのある概念構造を意味フレームと呼ぶが、その概念構造は明確な基本思想に基づいて構築されている。詳しくは、別報で述べるとして、ここでは、後で用いる意味フレームについて、簡単に説明する。

5. 動詞、形容詞の意味フレーム

図5に、後で述べる自然文の構造文を、表1にそのデータ構造を示す。意味フレームは図5で大きな[]で示した。例えば、意味フレームの"結婚する"は、図5と表1に示すように、PS8~PS9とMW31~MW42で表されている。//

印で単語を埋め込むMW(スロット)を示すが、単語が埋め込まれていない状態では、デフォルト値が設定されているので、

{誰かがいつかどこかで誰かと結婚する}

の意味になるが、もし、この意味フレームのMW38に(花子)が埋め込まれると、

{花子がいつかどこかで誰かと結婚する}

になる。___で示す部分は自然言語の原理Zによって表現禁止になるから、自然文では、次のようになり、意味が確定する。

{花子が結婚する}

表現されていない文字列はこのようにデフォルト値か、話し手が運用上の判断で表現禁止にしたかのどちらかである(この点に関しては別報で詳しく述べる予定である)が、本意味言語では、意味フレームは概念構造が明確に規定されていて、一切の曖昧性はない(自然言語の原理X)という立場をとっている(意味フレームの内部構造の詳細については、別報で述べる予定である)。意味フレームは概念辞書に登録されていて、意味解析のときに、動詞、形容詞などの単語に応じて、概念辞書から取り出されて、文章が表現する意味の概念構造を構築し、そのスロットに単語を埋め込んで意味を確定する。次に自然文がどのような意味構造を持っているかについて説明する。

6. 文章の意味の構造

{お金持ちの太郎と去年長崎で結婚した美しい花子は可愛い女の子の春子を生まれました}

という文章は図5に構造文で示すように、意味フレームの"美しい"、"お金持ち"、"可愛い"、"~は~である"、"結婚する"、"生む"が結合されて、文章全体の意味を表す概念構造を構築し、//で示されるスロットの中に、単語を埋め込んで、意味を確定する。これは意味解析である。意味解析には色々な方法があるが、上の文章の場合次のような方法で解析できる。

まず、形容詞の"美しい"の意味フレームを概念辞書から取り出す(図5のPS1~PS2とMW1~MW8)。この意味フレームのスロットはMW4とMM7であるから、"花子"をMW4に、MW7に"美し"と書き込む。同様にして、"お金持ち"の意味フレームに"太郎"と"お金持

ち”と, ”可愛い”の意味フレームに”女の子”と”可愛”を書き込む.”美しい”の”い”, ”可愛い”の”い”は助詞として, 取り扱っているので, 図5に示すように, ()の右脇やメンバjcsに書き込んで置く. この段階で意味解析がまだ終了していない文章は,

{太郎と去年長崎で結婚した花子は女の子の春子を生みました}

となるが, ”女の子の春子”は”春子は女の子である春子”であるから, ”～は～である”という意味フレームに”女の子”と”春子”を埋め込むと, まだ意味解析が完了していない部分は,

{太郎と去年長崎で結婚した花子は春子を生みました}

となる. つぎに, ”結婚する”の意味フレームを取り出し, これに図5に示すように, ”太郎”, ”去年”, ”長崎”を格助詞の”と”, ”で”をマークとして, 埋め込むと, まだ意味解析していない文は次のようになる.

{花子は春子を生みました}
”生む”の意味フレームに”花子”, ”春子”と

”生み”を埋め込み, 要素P Sの助詞のメンバに”ま”と”した”を埋め込めば, 上述の自然文の意味解析が完了する.

さらに複雑な文章もこのような手法で意味解析ができる(詳しくは別報で述べる)が, 自然文は基本的にこのような概念構造をもっている. この概念構造において, 基本的なルールで結合関係を変更することができる. 意味フレーム内でも可能であるが, 本論文では, 分かり易くするために, 意味フレーム間での結合関係の変更に限定して, 論ずることにする.

本意味言語では, 格結合(縦結合)と論理結合(横結合)は極めて近い関係にあり, 格結合から, 論理結合に, あるいは論理結合から格結合に容易に変更することができる.

次に格結合を論理結合に変更することによって, 仮説を生成できることを述べる.

7. 仮説の生成

”～ならば, ～である”というIF関係は次の

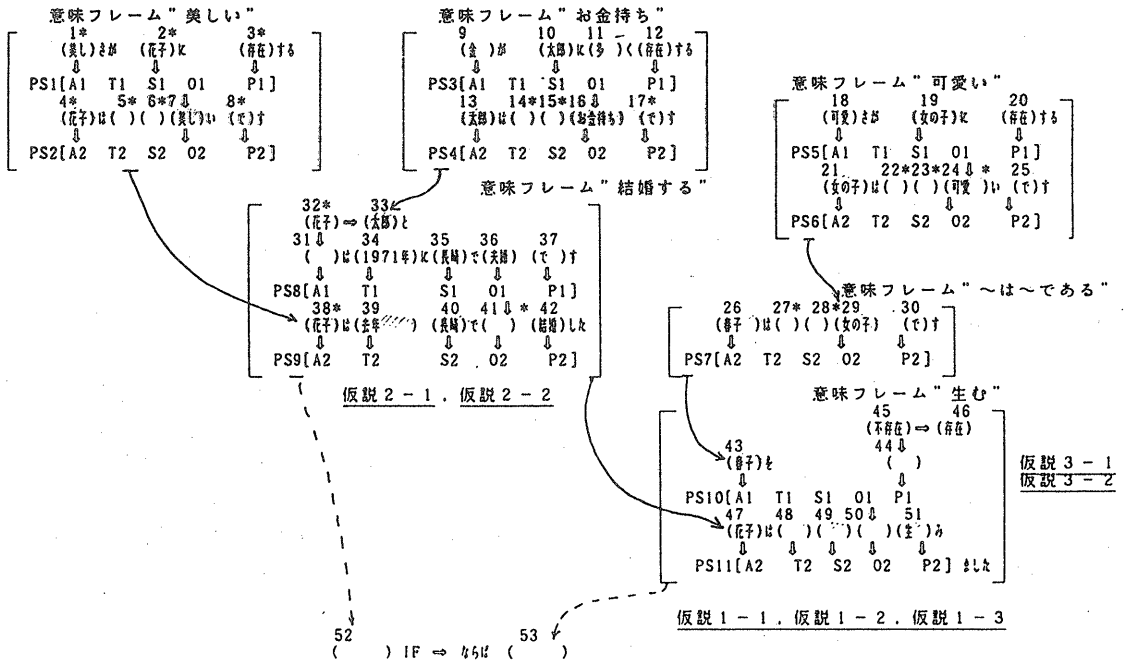


図5 【お金持ちの太郎と去年長崎で結婚した美しい花子は可愛い女の子の春子を生みました】の構造文と仮説の生成

ように要素MWを2個設け、それをIF関係で結合すればよい。つまり、

MW52 () IF ⇒ ならば MW53 ()

として、切り離れた意味フレームを結合する。

本意味言語では、基本的に下位-左側の方が意味の重要性が高い。図5の構造文において、最下位にある意味フレームは“生む”である。この意味フレームに結合している意味フレームのうち、下位-左側のルールを適応して、“結婚する”の意味フレームを切り離し、本体をIF論理関係の右側のMWに結合させ、切り離れた“結婚”の意味フレームを左側のMWに結合させると、

MW52 (PS9) IF ⇒ ならば MW53 (PS11)

となる。()内に結合したPSの番号を示した。この構造文から自然文を生成するには、構造文で指示されている順序に意味コードを並べ、これを日本語の文字列に変換すればよい。その結果次のような文が得られる。

{美しい花子が去年長崎でお金持ちの太郎と結婚すれば、花子は可愛い女の子の春子を生む}

となる。抽象化するために、時間、空間をデフォルト値にし、固有名詞を変数(変数名詞)の“X”、“Y”、“Z”に置き換えると、上の文は次のようになる。

{美しいXがお金持ちのYと結婚すれば、Xは可愛い女の子のZを生む}

... 仮説1-1

さらに条件を取り除いて、抽象化するために、“結婚する”と“生む”の意味フレームに結合されている意味フレームの結合関係を切り離し、自然文を生成すると、

{XがYと結婚するならば、XはZを生む}

... 仮説1-2

という仮説が生成できる。

表1にこの仮説のコンピュータ上のデータ構造を示す。意味フレームの“結婚する”以上の概念構造について同様な、処理を行うと、次のような仮説が得られる。

{Xが美しければ、Xはお金持ちのYと結婚する}

... 仮説2-1

意味フレームの“結婚する”に結合されている意味フレームの“お金持ち”を切り離すと、次のような仮説2-2が得られる。

{Xが美しければ、XはYと結婚する}

(要素 P S)														
no:	MK:	NTN:	L:	A:	T:	S:	O:	P:	X:	Y:	Z:	jntn:	jn:	jm:
8	100e		41	31	34	35	36	37						
9	e000		52	38	38	40	41	42						
10	100e			43				44						
11	e000		53	47	48	49	50	51				e	le	
(要素 MW)														
no:	BK:	MK:	LOG:	B:	N:	L:	MW	RP:	mw:	WD	jgb:	jcs:	jlg:	
31	001e	0001				8	32							
32	e001			33	31		38		0410					
33			32		0				0411			70		
34	e00e	0002			8		39							
35	e00e	0003			8		40						5	
36	e00e	0004			8				9720				5	
37	e00e	0005			8				1100		le			
38	100e	0001			9				0410			1		
39	100e	0002			9									
40	100e	0003			9							5		
41	00ee	0004			9	8								
42	100e	0005			9				9750		ab			
43	100e	0001			10				0412			4		
44	001e				10	45								
45	e001	0001		46	44				1000					
46	e000		45						1000					
47	100e	0001			11				0410			1		
48	100e	0002			11									
49	100e	0003			11									
50	00ee	0004			11	10								
51	100e	0005			11				2100					
52	00e0		5300		53		9						92	
53	00e0			52		11								

表1 {XがYと結婚すればXはZを生みます}のデータ構造

... 仮説2-2

“生む”の意味フレーム以上の概念構造について、同様な処理を行うと、次の仮説3-1が得られる。

{Xが可愛い女の子であれば、YはXを生む}

... 仮説3-1

同様に、意味フレームの“可愛い”を切り離せば次の仮説3-2が得られる

{Xが女の子であれば、YはXを生む}

... 仮説3-2

このような単純なルールを適用するだけで、これだけの仮説が生成できる。

8. 仮説の検証

上述の仮説はテキスト文から生成されたものであるが、この仮説が信頼性の高いものであるかは、その仮説が生成されたテキスト文以外でも成立するかどうかによって決まる。成立するかどうかは意味言語でのパターンマッチングによって行われる(意味言語でのパターンマッチングについては別報で詳しく述べる)。間違いなく常に成立するものはルールとなり、それをを用いて推論を行うことができるが、その仮説が生成されたテキスト文以外に成立しない仮説は信頼性に乏しいので棄て去られなければならない。

従って、その仮説の正当性はテキストの量や質

によって決まる。貧弱なテキストからは貧相な知識しか生まれないし、明確な論旨で豊富な内容を含んでいるテキストからは、質の高い知識が獲得される。

仮説の検証はテキストとのパターンマッチングによって行われるが、知識の信頼性が高い場合は、その知識から推論によって導かれた仮想的な世界と実世界との比較から、テキストの表現不足や誤りなどを指摘できる。

9. まとめ

人間は自然言語を用いて、質問応答、推論、知識獲得、学習、翻訳、物語理解などの知的活動を行っているため、自然言語の中にこれらの知的活動を支えているメカニズムが存在すると考えられる。そこで、本研究では自然言語の中からこれらのメカニズムを発掘する作業を行っている。

しかし、自然言語は個別言語の文法や話し手の運用上の判断や状況認識によって表現が著しく影響を受けるので、自然言語をそのまま用いるのではなく、人間の頭の中にある概念構造を記号化して取り出したものを意味言語と名付け、メタ言語として用い、上述の研究を行っている。

本論文では次の点が明らかになった。即ち、多層格-論理構造である意味言語の言語構造において下位-左側優先のルールに従って格結合をIF論理結合に変更し、さらに固有名詞を変数名詞に、時間、空間に関する単語をデフォルト値に変更することによって、自然文から、仮説文が自動的に生成できることが明らかになった。

このようにして獲得された仮説が成立するかどうかは、テキスト文とのパターンマッチングによって検証される。正当性を確保した仮説が知識になるが、この正当性はテキストの量や質によって決まる。貧弱なテキストからは貧相な知識しか生まれないし、明確な論理で豊富な内容を含んでいるテキストからは質の高い知識が獲得される。

本意味言語におけるパターンマッチングやそれによる知識獲得については別報で述べる予定である。

謝辞

長年にわたって多大のご支援を賜った九州産業大学工学部機械工学科楠本韶教授（元長崎大学工学部機械システム工学科教授）に心からお礼を申し上げます。また長期間にわたって研究プロジェクトに参加していただいた多くの方に感謝します。

参考文献

- (1) 人工知能学会編(1990): 人工知能ハンドブック, オーム社
- (2) 田中穂積, 辻井潤一編(1988): 自然言語理解, オーム社
- (3) Schank, R.C(1975): Conceptual Information Processing, North-Holland
- (4) 郡司隆男(1987): 自然言語の文法理論, 産業図書
- (5) Chomsky, N(1981): Lectures on Government and Binding, Foris, Dordrech. (安井稔, 原口庄輔監訳「統率. 束縛理論講義」研究社)