

意味言語の提案

安藤 司文

長崎大学工学部機械システム工学科

概要

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ものを意味言語となすけた。意味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影響を受けないので、自然言語の代わりに、意味言語で自然言語を研究することを提案した。そして自然言語と意味言語との関係、意味言語から見た自然言語の基本的な原理、自然言語の曖昧性について論じ、曖昧性は自然言語固有の特性でなく発生するものと考えて、その原因を挙げた。さらに意味言語の言語構造について簡単に説明し、自然文から意味言語への変換や意味言語から日本語と英語の自然文生成の基本的なプロセスを紹介した。

The Proposal of a Meaning-Language

Shimon Ando

Department of Mechanical Systems Engineering,
Faculty of Engineering, Nagasaki University

1-14, Bunkyou-machi, Nagasaki-city, Nagasaki, 852 Japan

abstract

The proposal of a meaning-language (imi-language) as Meta-language is made, where conceptional structure in a human brain is symbolized and taken out. Since the meaning language is not effected by the individual language or a speaker's judgement in performance of the language or perception of the situation, it is suggested to use the meaning language for studying a natural-language-processing instead of a natural language. In addition to the above, a relation between the natural language and the meaning language, a principle of the natural language, the fundamental linguistic structure of the meaning language, the transformation of the natural language to the meaning language and generation of Japanese and English natural language from the same meaning language are described in this paper.

1. はじめに

人間は自然言語を用いて思索を練り、自然言語を用いて互いに情報を伝達しているの、自然言語の中にこのような人間の知的な活動を支えているメカニズムが内在していると考えられる。従って、もしコンピュータに人間と同じ言語能力を持たせることができれば、人間とコンピュータとが自然言語で情報交換ができ、さらに人間と同じプロセスでコンピュータの知的能力を向上させることも期待できる。

しかし、自然言語には本質的に曖昧性があり、これをプログラミング言語のように用いることができないとされている。自然言語は個別言語の文法、話し手の運用上の判断、状況認識などによって表現が著しく影響を受け、多様な曖昧性が発生する。しかし、話し手の頭の中に存在する意味には一切の曖昧性はなく、その意味を表現する概念構造も簡潔明瞭で堅牢であると予想される。そこで、その概念構造を記号化して取り出してメタ言語として利用することを検討した。これは意味を正確に表現しているので意味言語となすけた。また、後述するように意味言語の一部の記号を個別言語の文字列に変換したものが自然言語になると考えられる。

意味言語は抽象的な言語で、個別言語のように特別な文字列を持っていない。また自然言語のように直接目で見ることができないが、前述のように意味言語が自然言語の母体であるとするれば、自然言語の中に意味言語の基本的な骨格が残っているはずであると考えられる。そこで、自然言語の中から意味言語の基本的な言語構造を発掘し、各種の方法で検証しながら、意味言語の全体の構造を明らかにすることにした。

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処理を行っているの、自然言語においてどのような方法でこのような知的処理能力が保証されるかを明らかにすることが、意味言語の言語構造を解明することになると考えられる。

従来の自然言語の研究は、文法理論、意味論、推論、機械翻訳、知識獲得など個別に研究されてきた(1)が、人間の知的処理は総合的に行われているの、自然言語を総合的に捉えなければならないと考えられる。

これまで多くの文法理論が提案されてきた(2~5)が、いずれも文法理論だけの立場で研究されてきた。例えば、Chomskyは統率・束縛理論(GB理論)(4)を、Gazdarらは一般化句構造文法(2)を、Bresnanらは語彙機能文法(2)を、Fillmoreは格文法(5)を提案しているが、その文法理論の中で翻訳のプロセスについて言及したものはない。また、人間は自然言語を用いて知識を獲得しているが、知識獲得についても一切明らかにされていない。自然言語の重要な役割は推論であり、知識獲得であるから、この機能を無視した文法は自然言語の解明の役に立たないと思われる。

Schankは、自然言語では意味が正確に表現されていないとしてもう一つ別の形式である概念依存構造(CD理論)(6,7)で自然文が表す意味を表現しようとしたが、本意味言語では自然言語だけで正確に意味が表現されているので、改めて意味を別な形式で表現し直す必要は一切ないという立場である。

最初に意味言語の基本的な文法理論について述べるべきであるが、分かりにくいと思われるので、本論文では、自然言語と意味言語との関係、意味言語から見た自然言語の原理、意味言語の基本的な言語構造、自然言語から意味言語への変換、意味言語からの自然言語の生成、自然言語における多様表現、コンピュータ上での意味言語のデータ構造などごく基本的な枠組について報告する。

2. 自然言語と意味言語との関係

2.1 意味言語による情報の伝達

話し手が自然言語を表出するときには、まず、図1に示すように、話し手の頭の中に表現したいと考

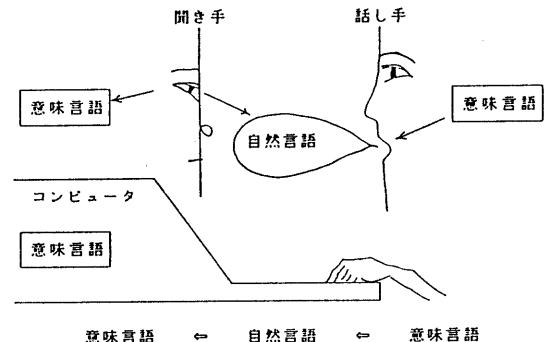


図1 意味言語と自然言語との関係

えている概念構造（意味言語）が存在しなければならぬ。話し手は聞き手の判断能力やそのときの状況を認識しながら、頭の中にある概念構造を加工し、個別言語の文字に変換して、表出するが、これが自然言語であると考えられる。聞き手はその自然言語から、話し手が表現しようとしている概念構造（意味言語）と同じものを自分の頭の中にある概念構造から取り出すか、なければ作り出す。話し手の概念構造（意味言語）と同じ概念構造（意味言語）が取り出されたときに情報伝達が完結する。つまり、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、自然言語には直接関係がない。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のすぎない。前述のように自然言語は話し手の判断によって著しく変化し、また情報伝達に関しても決定権をもっていない。ここで、意味言語から見た自然言語の基本的な原理について述べる。

2. 2 自然言語の原理

自然言語の原理として、次の三つを挙げる。

- 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：必要性のない単語および助詞はできるだけ表現を省略し、表現されている単語およ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。

話し手の頭の中にはまず明確な概念構造がなければならぬ。図2には、例として、

{太郎は今日学校で花子に本を与えた}

という概念構造を示した（この概念構造については後で詳しくのべる）。この概念構造から自然文を生成するには、同図で示した、自然文生成経路に沿って、単語や助詞を個別言語特有の文字列に変換すればよい。日本語の文字列で表出したのが、図2に示す次の文である。

{太郎は花子に本を与えた}

聞き手はこの自然文（自然言語）から、自分の頭の中に、話し手と同じ概念構造を形成しようとす

る。図2に示すように、聞き手の頭の中に、話し手とはほぼ同じような概念構造が取り出されると、情報伝達が完結することになる。

話し手は太郎が花子に本を与えた時間の”今日”や場所の”学校”について知っているのに何も言及していないので、聞き手の頭の中にある概念構造には、同図に示すように、”今日”，”学校”の単語が埋め込まれていない。聞き手にとってはこの点が曖昧である。

図2に示した例で、自然言語の原理について説明する。原理Xは自然言語、即ち自然言語の母体である意味言語、の概念構造（言語構造）には一切の曖昧性はないことを表している。つまり、この概念構造でこの自然文で表現される意味は正確に表現されていて、曖昧性はないと考える。

聞き手は話し手が表出した自然言語から話し手が表現しようとした概念構造と同じものを多くの概念構造から選別するか、なければ構築しなければならないが、この場合、聞き手が”与える”という意味フレーム（意味フレームについては後述する）を取り出して、そのスロットに”太郎”，”花子””本”などの単語を、格助詞の”は”，”に”，”を”の指示に従って埋め込んで意味を確定する。原理Yはこのことを表している。

聞き手がその概念構造を正確に再構築するため、話し手は必要最小限の内容しか表現すべきでないというのが原理Zである。この原理Zは徹底的に実行されるので、聞き手が必要とするもの以外は表現が省略される。話し手の概念構造のMW3とMW12に”太郎”が、MW7とMW5に”花子”が二回埋め込まれているがMW7とMW12しか表現されていない（表現禁止^①印で示す）。これはMW3とMW5の内容は聞き手にとって容易にわかるので、表現する必要がないとして原理Zによって表現が禁止になったと考えられる。

”今日”と”学校”は聞き手によっては曖昧であると先に述べた。本研究では、自然言語に曖昧性はなく、曖昧性は発生するものであると考えて、曖昧性発生原因として次の四つを挙げた。

2. 3 曖昧性発生原因

- 発生原因1：話し手があいまいに表現する場合。
つまり、話し手が話題として取り上げる価値がないと判断した場合や話

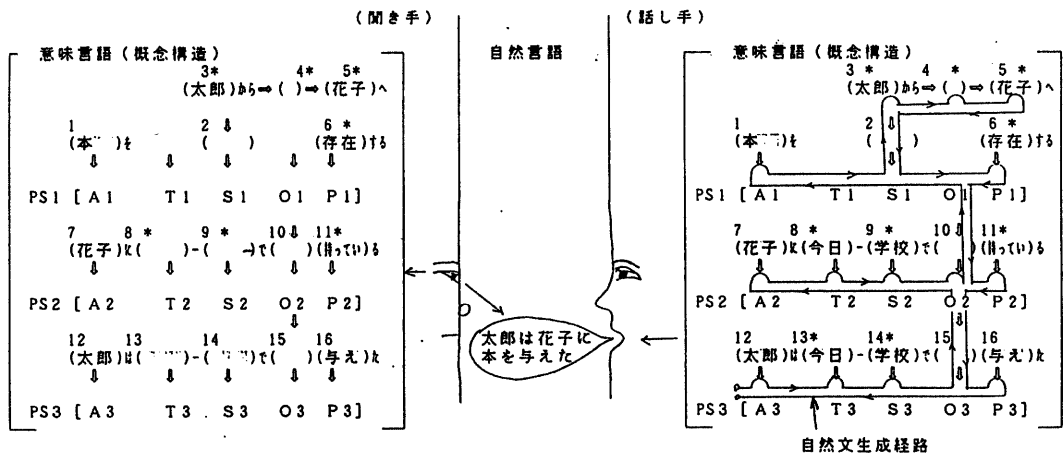


図2 意味言語による情報の伝達

し手にとって都合が悪いので、触れたくない場合や内容が間違っている場合など。

- 発生原因2 自然文は原理Zによって unnecessaryな単語や助詞は徹底的にその表現が省略されるが、その必要性は話し手によって一方的に決められるので、その判断が適切でないか、聞き手が必要とする内容が表現されていない場合
- 発生原因3 単語、助詞の表現が省略されすぎていて文法的に不適切な場合
- 発生原因4 ことば遊びなどで二重の意味を持たせる場合

図2で話し手が“今日学校で”を表現しなかったのは、話す必要がないと判断した為であったとするならば、これは発生原因1によることになる。このような曖昧性の発生原因を設定して、曖昧性をいくつかの種類に分類し、解消すべき曖昧性か、解消すべきでない曖昧性かを峻別しようとしている。この点に関しては別報で述べる予定である。

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を生成させ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だけしかない。動詞、形容詞はあるまとまりのある概念構造（意味フレーム）を代表する単語で、意味解析のときにしか用いない。主語、目的語、補語などは概念構造のある特定の位置の単語に付けられた役割で、本文法ではほとんどどの役割も持っていない。単語以外は全て助詞で、そのため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。

要素MW (META WORDの略)には概念を表す単語が格納される。本論文では、分かり易くするために、図2で示したように、単語を日本語で表現しているが、コンピュータの中では、後出の表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MWに格納される。要素MWを()の記号で表す。他の要素MWや後述の要素PSと結合するために結合手をもっているが、これを⇒や↓で表す。要素MWは、図3に示すように、上下2本と左右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をもっているが、本論文ではこの点に触れないので省略する。“あの”などの冠詞も助詞として取り扱

っているが、半角で () の左側に示し、格助詞は () の右側に示す。従って
 ”あの(太郎)が”
 のように表記される。

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基本文を表す要素として 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。即ち

- A格：主格句 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と結合する格で、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。つまり要素PSは、who, when, where, what, howを示す単語を格納する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IH5W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

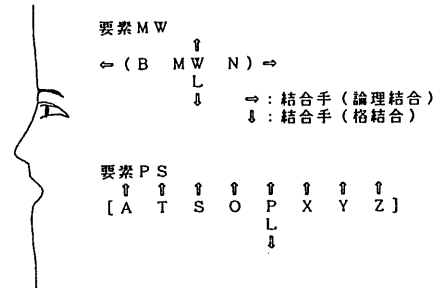


図3 意味言語を構成する要素、MWとPS

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。これについては別報で詳しく述べる。

本意味言語では格はA、T、S、O、Pの五つの格と、補助的なX、Y、Zの格だけである。

Fillmoreは多くの格を考えているが(5)、これらの格は、PSやMWが多層に組合わされたときに表れる特定のMWがそれに相当する。この点に関しては別報で詳しく述べる予定である。

結局、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本の結合手を持っている。即ち、人間の頭の中にある概念構造は図4に示すように、上下左右の4本の結合手を持った要素MWと9本の結合手をもった要素PSによって構築される。

3.3 構造文とデータ文

前述のように、要素MW () と要素PS [] を縦(↓格結合)と横(⇒論理結合)に結合させて、人間の頭の中にある概念構造を表記するが、これは意味言語の言語構造を図式的に分かり易く表現しているので、構造文と呼ぶことにする。コンピュータ上では、後出の表1に示すようなデータ構造で自然文が表す意味が表現される。これをデータ文と呼ぶことにする。

図4の(a)で示されて構造文の結合関係をデータ文で示すと(b)のようになる。データ文は要素PSと要素MWに分かれている。

PS1のA格はMW1と結合しているので、P

即ち, MW 1 2, MW 1 3, MW 1 4を埋め込み禁止にする. MW 2 3, MW 2 4, MW 2 5に単語が埋め込まれた場合, 直ちにその単語をコピーするために, 相手先のMWの番号をメンバーPに記録しておく(表1). このような対策を立てておくと, 単語を埋め込むスロットが少なくなる. この文章の場合, 格助詞の"は", "に", "で", "を"をマーカにすれば, 図5, 表1のように単語を各MWに埋め込むことができ, 意味解析は終了する.

この場合も係受けの解析は一切行わない. 上述の文章はいずれも簡単であるが, 複雑な文章でも同様な方法で行うことができる.

上の文はコンピュータ上では表1に示すようなデータ構造(データ文)で表される.

5. 意味言語からの多様な自然文の生成

表1で示したデータ文は生のデータ群であるが, 文の意味をコード番号やデータで記述されているため, 個別言語にとらわれない普遍的な言語である. この普遍的な言語からは, 各言語固有の辞書を用いると, 英語, 日本語などの自然文が容易に生成できる.

機械翻訳の分野で存在が予想されている中間言語がこれに相当すると考えられる(1).

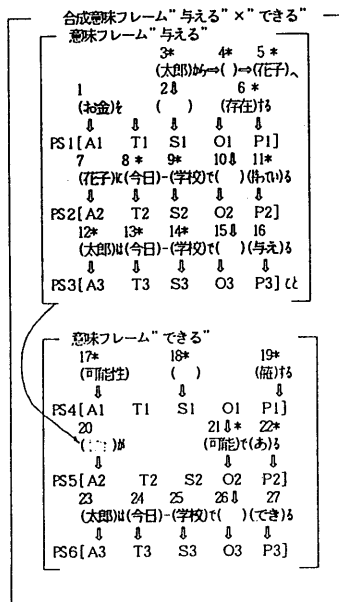


図5 (太郎は今日学校で花子にお金を与えることができる)の構造文

詳しくは別報の"普遍文法"で述べる予定であるが, 個別言語には固有の標準語順がある. 自然言語生成経路をその標準語順をベースにして指定し, その経路を辿りながら, 表現が許可されているMWの単語や助詞の意味コードを単語辞書を用いて, 文字列に変換すれば容易に自然文が生成できる. 英語と日本語では助詞が異なったり, 習慣的に表現する内容が異なることがある. そのためデータ文を完全に同じにすることは無理であるが, 基本的に英語, 日本語ともにほぼ同じデータ構造(データ文)になる.

図2の自然文は日本語の標準語順による自然言語生成経路に沿って生成された文である. 同図において, 同じ単語が埋め込まれているMWで, 表現許可のMWを変えると, いろいろな語順の自然文が得られる. 例えば, MW 1 2とMW 1 4を表現禁止にして, その代わりにMW 3とMW 9を表現許可にすると, 同じ経路で,

{今日花子に学校で本を太郎から与えた}
という文が生成できる. 次に示す英文も簡単に生成できる.

{Taro gave Hanako a book at school today}
この意味言語において, 表現禁止をMW 7とし, 表現許可をMW 5とすると

{Taro gave a book to Hanako at school today
になる." gave(過去)"や"a book"の"a"

(要素PS)		LI AI TI SI O; PI XI YI ZIjntlnlnlnjstl												
noI	MXI	HWI												
11100e1			101	11	21	1	61							
21			151	71	81	91	101	111						
31e00e1			201	121	131	141	151	161						311
41100e1			211	171	181	191								
51			261	201		121	221							
61e00e1			1231	241	251	261	271							
(要素MW)		noI BXI MXI LOCi BI NI L MW RP mw WD jgb jcs j g jostl												
noI	BXI	MXI	LOCi	BI	NI	L MW RP mw	WD jgb jcs j g jostl							
11100e100011							16b121							41
21001e100031														
31c0011				41	21	121						e4511		101
41e0001				31	51									
51c0001				41								e4611		81
61e00e100051												10001	ab1	
71100e100011												e4611		31
81e00e100021												d4201		
91e00e100031												c6af1		51
10100e100041														
111e00e100051												30111		481
121e00e100011												e4511		11
131e00e100021												d4201		
141e00e100031												c6511		51
15100e100041														
161100e100051												131021		481
171e00e100011												154001		
181e00e100031														
191e00e100051												10001	ab1	
20100e100011														21
21100e100041												154101		51
221e00e100051												111001		481
231100e100011												e4511		11
241100e100021												d4201		
251100e100031												c6af1		51
26100e100041														
271100e100051												154011		481

表1 (太郎は今日学校で花子にお金を与えることができる)のデータ構造(データ文)

については、別報で詳しく述べる予定である。

表1から次のような日本語や英文を生成することができる。

{太郎は今日学校で花子にお金を与えることができる}

{太郎は今日学校で花子にお金を与えられる}

{太郎が今日学校で花子にお金を与えることは可能である}

{Taro can give Hanako money at school today}

{Taro is able to give Hanako money at school today}

{It is possible that Taro gives Hanako money at school today}

{It is possible for Taro to give money to Hanako at school today}

などの多様な表現の文章が一つの意味言語から生成できる。

6. まとめ

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ものを意味言語となすけた。意味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、表現が影響を受けないので、自然言語の代わりに、意味言語を用いて自然言語を研究することを提案している。

本論文で、自然言語と意味言語との関係、意味言語から見た自然言語の基本的な原理、自然言語の曖昧性などについて論じ、曖昧性は自然言語固有の特性でなく、曖昧性は発生するものと考え、その発生原因を挙げた。さらに意味言語の言語構造について簡単に説明し、自然文から意味言語の変換や意味言語から日本語と英語の自然文生成の基本的なプロセスを紹介した。

この意味言語は機械翻訳で存在が予想されている中間言語に相当するものと考えている。本論文では意味言語のごく基本的な枠組について述べたが、意味言語の言語構造、意味概念の構築、意味解析、多様な自然文の生成、多言語間機械翻訳などの詳細については別の機会に述べる予定である。

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、長期間にわたって本研究のプロジェクトに参加していただいた多くの方々に感謝します。

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) Peters, S(1985):Lectures on Contemporary Syntactic Theories, (郡司隆男, 田窪行則, 石川彰訳「現代の文法理論」産業図書)
- (3) 郡司隆男(1987):自然言語の文法理論, 産業図書
- (4) Chomsky, N(1981):Lectures on Government and Binding, Foris, Dordrecht. (安井稔, 原口庄輔監訳「統率・束縛理論講義」研究社)
- (5) Fillmore, C, J(1975):Toward a Modern Theory of Case & Articles, (田中春美, 舟城道雄訳「格文法の原理-言語の意味と構造」三省堂)
- (6) 田中穂積, 辻井潤一共編(1988):自然言語理解, オーム社
- (7) Schank, R, C(1975):Conceptual Information Processing, North-Holland