

意味言語による多言語間機械翻訳

英語⇔マレー語⇔日本語の場合

安藤 司文*

四元 健二

*敬和学園大学

住友金属工業株式会社

要旨

人間の頭の中に形成されてい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て、これによって自然言語の文法理論の構築を行っている。

人間はこの意味言語を用いて、質問応答、知識獲得、推論、学習、物語理解、翻訳などを行っているので、この意味言語を用いて、このような人間の知的な機能をコンピュータ上で実現することを試みている。

本論文では、英文を一旦意味言語に変換し、それから英文に再変換させたり、マレー語と日本語の自然文を生成することによって多言語間の翻訳を行っている。このことから、意味言語によって同時に多言語間の機械翻訳が可能であることを示す。

The Machine-Translation between Multi-Languages Based on Meaning Language

Shimon Ando*

Kenji Yotsumoto

*Keiwa College

Sumitomo Metal Industrial, LTD

Abstract

"Meaning language" is the name of the conceptual structure existing in a human brain which is taken out after being symbolized. Using this meaning language, we tried to construct a grammatical theory of natural languages. Human beings use the meaning language to answer question, acquire knowledge, reason, learn, understand stories, and translate languages. We thus tried to recreate such mental activities using computers that also speak this meaning language.

This paper reports on our experiment on translation between several languages. We first translated English sentences into the meaning language. The resulting sentences were then translated back into English, or used to form natural sentences in Japanese and Malay. Through these experiments, we demonstrated that the meaning language enables simultaneous machine translations between several languages.

1. はじめに

長年の間、溶接ロボットや塗装ロボットなどを中心に知能的な機械の研究開発に従事してきた。これらのロボットや知能機械は開発当時は一応知能的な機械として認められていたが、たいした知能をもっているわけではなかった(4)。

産業用知能ロボットの開発が一段落した機会に、人間と知的な会話ができ、しかも人間のように自分自身で知識を獲得できる本格的な知能ロボットを開発することができないかどうかを検討し始めたが、まず最初にいったい“知能とは何か?”という問題に直面した。“知能とは何か”については分からなかったが、しかし、人間は自然言語を用いて質問応答、知識獲得、学習、物語理解、翻訳などの知的な活動を行っているので、自然言語を研究すれば知能について解明できると考えた。

自然言語の研究は文法理論の分野で昔から盛んに行われてきたが(1)～(3)、これまでの文法理論では主語、0役割、動作主、道具格などの概念で自然言語を解明しようとしてきた。しかし、人間の知的なプロセスを解明しようとするには、このような概念では不十分であると考えて、新しい概念に基づく文法理論を考案することにした。人間が表出する自然文である文字列は話し手が使用する個別言語に決定的に左右されることと、話し手の状況認識や運用上の判断によって著しく影響を受けるので、そのような不安定な材料をベースにして自然言語の研究を行うことは不可能であると判断した。話し手が自然文を表出するためには、話し手の頭の中に明確な概念構造が存在するはずであるので、それを記号化して取り出したものを意味言語となづけ、これをベースに自然言語の研究を行うことにした。

本意味言語では概念構造と自然文の単語、助詞、記号などとは明確に対応しているので、概念構造の変化とそれに伴う自然文の変化を単語、助詞、記号のレベルで正確に追跡できる。これまで、日本語の使役文や受け身文を例にとって日本語の言語構造を明らかにしたり(12)、また、ある一

つの自然文から仮説文を含めた多種多様な自然文が生成できるプロセスを説明し(13)、自然言語のなかに我々人間の知的な活動を支援するメカニズムがあるとことを明らかにした。このように我々人間はこの意味言語を用いて前述のような知的な活動を行っているので、本研究ではコンピュータ上で上述のような機能を検証しながら、文法理論の構築を行っている(5)～(13)。

また、質問応答は人間の知的な機能にとっては基本的な機能であるが、この質問応答もコンピュータ上で正確にしかも容易に実現できる見通しが得られた(8)。

意味言語は人類にとって普遍的な概念構造で記述されるので、同時に多言語間の翻訳を行うことは原理的に比較的容易である。

本論文では形態解析によって、英文をある種の言語構造情報に変換し、それをさらに意味解析によって一旦意味言語に変換する。この普遍的な言語である意味言語から英語に再変換したり、マレー語や日本語に変換することによって、多言語へ同時機械翻訳を行う。これらのプロセスを紹介する

意味言語についてはすでに紹介しているが、ここでごく簡単に意味言語の概要について述べる。

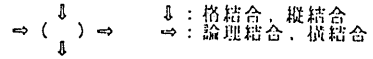
2. 意味言語の概要

意味言語を表現する記号として、概念を表す語要素 MW (META WORDの略)と基本的な意味単位を表す文要素 P S (PRIMITIVE SENTENCE)を設定している。MWを()、P Sを[]の記号で表す。単語はある概念を表現するのでMWに単語を埋め込めば、MWは概念を表現することになる。文要素のP SにはA格(主格, AGENT CASE)、P格(述語格, PREDICATE CASE)、O格(対象格, OBJECT CASE)、S格(空間格, SPACE CASE)、T格(時間格, TIME CASE)と補助的なX格(補助格, AUXILIARY CASE)、Y格(応答格, YES-NO CASE)、Z格(全体格, ZENTAI CASE)から構成されている。MW格は図1(a)に示すように4本

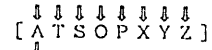
PSは(b)に示すように9本の結合用の手を持っている。横結合(論理結合)を⇒で、縦結合(格結合)を↓で示す。PSのA, P, O, S, Tの5個の格にMWを縦結合させ、さらにMWを介して, IFという論理関係で結合すると, (c)に示すような概念構造が形成される。これは(who), (when), (where), (what), (how), (why), 即ち, 1H5Wを示す。本意味言語ではこれを基本意味単位とする。このPSのP格に結合されているMWに"ある(存在)", "ある(状態)", "する(行為)"などの単語を埋め込むと, "～がある"(PS-EX), "～である"(PS-IS), "～をする"(PS-DO)という3種類の基本意味単位が得られる。この基本意味単位が縦横に組み合わされて複雑な概念構造が表現される。我々人間は日常生活の中であるまとまった概念構造を頻繁に用いているが, これに動詞, 形容詞を割り当てている。これを意味フレームと呼ぶことにするが, 意味フレームにはスロットがあり, これに単語を埋め込めば意味を特定することができる。(e)に"与える"の意味フレームを示す。同図に示すようにスロットに"太郎", "今日", "学校", "花子", "金"を埋め込むと{太郎は今日学校で花子に金を与えた}という自然文が表現する概念構造が得られる。この意味フレームはPS1, PS2, PS3の3階層で構成されており, PS1は"金の太郎から花子への移動", PS2は"花子はそのような状態である"こと, PS3は"太郎がそのような状態にした"ことを表している。この意味フレームには同図に示すように, 同じ単語が何回も埋め込まれているが, <必要性のない単語は表現を禁止にする>という自然言語の原理Z(6)によって, 最重要の単語を一つだけ残し, 他は表現が禁止(*印で示す)になる。一般に意味は下位にあるほど意味の優先度が高い(下位優先ルール)ので, 下位にある単語だけが表現可能になる。自然文の概念構造は(f)に示すように, 意味フレームを中心にして, MWやPSによって縦横に結合された構造で表現される。

3. 自然言語⇒意味言語⇒自然言語への変換

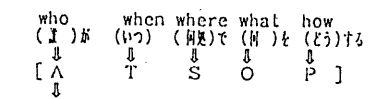
自然言語は意味言語へ, さらにその意味言語は自然言語へ再変換されるが, そのプロセスの概要を図2に示す。英語から意味言語へ変換し, それをさらに英語に変換すれば, 表面的に何の変哲も



(a) 構成要素 MW

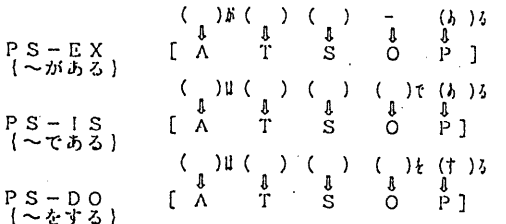


(b) 構成要素 PS

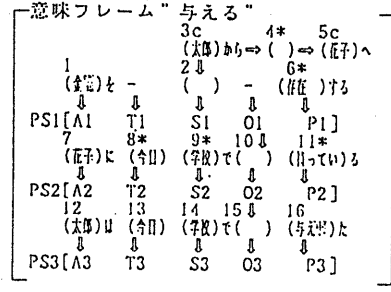


why (理由) IF ⇒ 4H5W ()

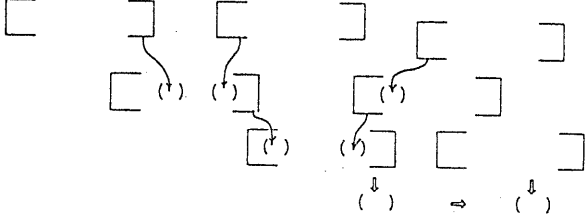
(c) 基本意味単位



(d) 三種類の基本意味単位



(e) "与える"の意味フレーム (太郎は今日学校で花子に金を与えた)



(f) 文章の意味構造

図1 意味言語の言語構造

ない処理に見えるが、コンピュータ上ではこの意味言語（PSデータ、MWデータ）で質問応答、知識獲得、推論、物語理解、学習などを行うので、自然言語から意味言語に正確に変換することは極めて重要である。多言語間の翻訳は、意味言語から他の多くの言語に変換することであるから、本意味言語の研究では、多言語間の機械翻訳は全体の研究の一部として取り扱うことができる。自然文が入力されると、形態解析文法によって、その文字列を各種の辞書上で検索しながら、言語構造情報に分解して、それをWSテーブルとして格納する。意味解析を行うときに、処理の途中で各種の情報が必要になるので、情報を書いたり消したりできる作業用テーブル（MKテーブル）をWSテーブルから作成する。このMKテーブルのデータに対して意味解析文法を適用しながら、PSデータとMWデータ（意味言語）の形で、自然文が表現する概念構造をコンピュータ上で表現する。自然文は文生成辞書を検索しながら、この意味言語のPSデータとMWデータから自然文生成文法によって個別言語固有の文字列に変換して、生成されることになる。各国は固有の文字列をもち、しかも自然文の表現形式は個別言語固有の文法に支配される部分が多い。従って、入力自然文や出力自然文に関係の深い単語辞書、形態解析辞書、文生成辞書、形態解析文法、自然文生成文法には個別言語固有の情報を記載しなければならず、各言語別の辞書や文法が必要になる。意味言語は世界共通の概念構造で表現されるので、これに関連した辞書、即ち、後述の意味フレームが記載されている概念辞書や意味解析文法には普遍的に記述できる部分は多くなる。しかし、完全にそのようにすることができない。特に、助詞や記号は本意味言語に於ては、概念構造に用いられる一種のマークで、その役割は個別言語によって左右されるところがあり、世界共通にすることは難しい。そのために意味フレームに於ても個別言語に依存する部分が残る。辞書、文法に関しては、別の機会に改めて詳述する予定である。

4. 英文の形態解析

次に示す英文

{ We visited Mito-Park that is famous for its plum-blossoms. }

の場合は次の様にして形態素解析を行う。まず、形態解析文法の指示によって、単語辞書を検索して、“visited”は規則動詞“visit”の過去形で、語尾変化する助詞（本意味言語では助詞をして扱っている）が“ed”であることを検出して、形態情報KTや論理情報LOGや時制否定情報NTNに（KT/522）、（NTN/2）のようにデータを書き込む。単語辞書からその他の情報を表1に示すように書き込む。同様に、“is”が“be”動詞の三人称単数現在であることを検出して、（KT/311）と（NTN/0）を書き込む。“plum-blossoms”は名詞の“plum-blossom”に“s”が付いて、複数になったものであること検出して、（KT/987）と（LOG/1）を書き込む。この形態情報KTを自然文生成のときに使用すると、同じ文が生成できる。ここでは“that”は関係代名詞の“that”（LS/81）であることが検出されたとする（この点に関しては別の機会に報告する）

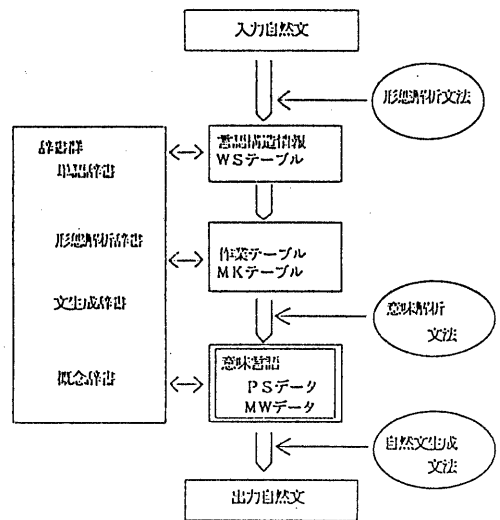


図2 自然言語 → 意味言語 → 自然言語への変換

は、別の機会に報告する)。

その他の情報は表1に示すようになる。品詞情報はLSに格納され、"11"が名詞、"12"が動詞、"17"が形容詞、"81"が関係代名詞、"53"が論理助詞、"61"が冠詞、"9a"が記号である。WDにはSPELLLに示した単語の意味コードを16進数の4桁で示している。JOは助詞の意味コードである。PTNに意味フレームの番号を示す。

5. 意味解析

WSテーブルで示された言語情報から意味解析文法によって、意味言語に変換するが、このとき、これらの言語情報を処理しなければならないので、WSテーブルの一部をコピーして、表1のSTEP1で示すようなMKテーブルを作成する。MKテーブルにはWSテーブルの番号WSNOが書き込まれているので、自由にWSテーブルに飛ぶことができる。またWSテーブルからはWDの単語の意味コードによって、自由に単語辞書に飛び、不足している情報を取り出すことができる。

この意味解析では、できるところから始める。どのような戦略で行うかは意味解析のポイントであるが、本論文の英文はごく簡単な文であるから、次のようにする。

"its"と"plum-blossom"は関係があるから、MW1を表3に示すように設けて、MW1の冠詞jarに"its"のコード番号"48"を"plum-blossom"のコード番号"ba90"をWDに書き込み、MKテーブルには、PSMWに"MW"を、NOにはその番号の"1"を書き込む。次に関係代名詞(LS/81)以降で解析できるものを捜す。LS/12(動詞)+LS/17(形容詞)となっているので、形容詞"famous"の意味フレームPS1-PS3, MW2~MW14(PTN/9)を概念辞書DICT-IMIから取り出し、表3に書き込む。この意味フレームには、スロットがあり、MW11に"1103"(be)を、MW12には"6c21"(famous)を書き込む。図式的に示すと図3のよ

表1 WSテーブル

{ We visited Mito-Park that is famous for its plum-blossoms. }

no	WD	JO	LS1	LS3	LS4	KT	PTN	NTN	LOC	SPELL
1	1903	1	011	18	10	1	01	01	01	we
2	1361	1	0112	10	10	522	71	21	01	visit
3	c611	1	0111	13	10	1	01	01	01	Mito-park
4	01	5c	181	10	10	1	01	01	01	that
5	1103	1	012	10	10	311	21	01	01	be
6	6c21	1	017	10	10	51	91	01	01	famous
7	01	13	53	10	10	1	01	01	01	for
8	01	48	61	10	10	1	01	01	01	its
9	ba90	1	011	10	10	978	01	01	11	plum-blossom
10	01	c1	9a	10	10	1	01	01	01	.

表2 MKテーブル (作業テーブル)

STEP 1

no	MK	PSMW	NO	LS1	LS3	LS4	JO	WSNO	SPELL
1	11	221	01	01	111	81	01	01	11 we
2	21	221	01	01	121	01	01	01	21 visit
3	31	221	01	01	111	31	01	01	31 Mito-Park
4	41	221	01	01	811	01	01	5c1	41 that
5	51	221	01	01	121	01	01	01	51 be
6	61	221	01	01	171	01	01	01	61 famous
7	71	221	01	01	531	01	01	131	71 for
8	81	221	01	01	611	01	01	481	81 its
9	91	221	01	01	111	01	01	01	91 plum-blossom
10	881	01	01	9a1	01	01	01	c1	101 .

STEP 2

no	MK	PSMW	NO	LS1	LS3	LS4	JO	WSNO	SPELL
1	11	221	01	01	111	81	01	01	11 we
2	21	221	01	01	121	01	01	01	21 visit
3	31	221	01	01	111	31	01	01	31 Mito-Park
4	41	221	01	01	811	01	01	5c1	41 that
5	51	221	01	01	121	01	01	01	51 be
6	61	221	01	01	171	01	01	01	61 famous
7	71	221	01	01	531	01	01	131	71 for
8	81	221	MW1	11	111	01	01	01	91 MW1
9	881	01	01	9a1	01	01	01	c1	101 .

STEP 3

no	MK	PSMW	NO	LS1	LS3	LS4	JO	WSNO	SPELL
1	11	221	01	01	111	81	01	01	11 we
2	21	221	01	01	121	01	01	01	21 visit
3	31	221	01	01	111	31	01	01	31 Mito-Park
4	41	221	01	01	811	01	01	5c1	41 that
5	51	221	PS1	31	121	01	01	01	51 PS3
6	61	221	01	01	531	01	01	131	71 for
7	71	221	MW1	11	111	01	01	01	91 MW1
8	881	01	01	9a1	01	01	01	c1	101 .

STEP 4

no	MK	PSMW	NO	LS1	LS3	LS4	JO	WSNO	SPELL
1	11	221	01	01	111	81	01	01	11 we
2	21	221	01	01	121	01	01	01	21 visit
3	31	221	01	01	111	31	01	01	31 Mito-Park
4	41	221	01	01	811	01	01	5c1	41 that
5	51	221	PS1	31	121	01	01	01	51 PS3
6	61	881	01	01	9a1	01	01	c1	101 .

STEP 5

no	MK	PSMW	NO	LS1	LS3	LS4	JO	WSNO	SPELL
1	11	221	01	01	111	81	01	01	11 we
2	21	221	01	01	121	01	01	01	21 visit
3	31	221	01	01	111	31	01	01	31 Mito-Park
4	881	01	01	9a1	01	01	01	c1	101 .

STEP 6

no	MK	PSMW	NO	LS1	LS3	LS4	JO	WSNO	SPELL
1	11	881	PS1	51	111	81	01	01	11 PS5

うになる。これで“be”と“famous”が解析でき、意味フレームの中に取り込んだので、それを示すためにPSMWにPSを、NOに番号の“3”を書き込み、“be”と“famous”をPS3に置き換えると、STEP3のようになる。この段階でLS/53(～なので～の含意の論理助詞“FOR”)でPS3とMW1でつながっているので、図3に示すような論理関係を構成する。即ち、PS3の下にW15を設け、MW1と“FOR”の論理関係で横結合させる。

これで同図に示すような合成意味フレーム“famous×for”を作る。この合成意味フレームの代表をPS3とすると、この段階でMKテーブルはSTEP4のようになる。関係代名詞“that”の前にある単語“Mito-Park”をこの合成意味フレーム埋め込む。この合成意味フレームに埋め込み探索経路を設定して、埋め込み可能な最初のスロットMW10に埋め込むが、表現禁止(*印で示す)にする。未解析の部分はSTEP5に示すようになる。

表3 意味言語のデータ構造

(a) PSデータ構造

no	MK	INTN	KT	A	P	O	S	T	L	jk1	jntn	jn	jm	jost	jk2
1	100e	0	0	2	3	0	4	0	7	0	0	0	0	0	0
2	e	0	0	5	6	7	8	9	12	0	0	0	0	0	0
3	e00	0	0	10	11	12	13	14	15	0	0	0	5c	0	0
4	100e	0	0	16	17	0	18	0	21	0	0	0	0	0	0
5	e000	2	0	19	20	21	22	23	0	0	0	0	0	0	e

(b) MWデータ構造

no	MK	BK	LOG	KT	MW	L	N	B	RP	mw	WD	CNC	jar	jhd	jul	jls	jlg	jcs	Jos	jdx
1	0	1000	0	978	0	0	0	15	0	0	ba90	0	48	0	0	0	0	0	0	0
2	1	e00e	0	0	0	1	0	10	0	c611	0	0	0	0	0	0	0	0	0	0
3	5	e00e	0	0	0	1	0	0	0	1000	0	0	0	0	0	0	0	0	0	0
4	3	e00e	0	0	0	1	0	0	5	9500	0	0	0	0	0	0	0	0	0	0
5	1	e00e	0	0	0	2	0	0	0	9500	0	0	0	0	0	0	0	0	0	0
6	5	e00e	0	0	0	2	0	0	0	3941	0	0	0	0	0	0	0	0	0	0
7	4	e0ee	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	3	e00e	0	0	0	2	0	0	13	0	0	0	0	0	0	0	0	0	0	0
9	2	e00e	0	0	0	2	0	0	14	0	0	0	0	0	0	0	0	0	0	0
10	1	e00e	0	0	0	3	0	0	0	c611	0	0	0	0	0	0	0	0	0	0
11	5	100e	0	311	0	3	0	0	0	1103	0	0	0	0	0	0	0	0	0	0
12	4	10ee	0	0	2	3	0	0	0	6c21	0	0	0	0	0	0	0	0	0	0
13	3	100e	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	2	100e	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	00e1	9000	0	3	18	1	0	0	0	0	0	0	0	0	0	0	13	0	0
16	1	e00e	0	0	0	4	0	0	19	0	9031	0	0	0	0	0	0	0	0	0
17	5	e00e	0	0	0	4	0	0	0	1000	0	0	0	0	0	0	0	0	0	0
18	3	101e	0	0	15	4	0	0	0	10	6c11	0	0	0	0	0	0	0	0	0
19	1	100e	0	0	0	5	0	0	0	0	9031	0	0	0	0	0	0	0	0	0
20	5	1000	0	522	0	5	0	0	0	0	3611	0	0	0	0	0	0	0	0	0
21	4	00ee	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	3	100e	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	2	100e	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0

“visit”の意味フレーム(PS3~PS4, MW16~23)を概念辞書から取り出して、表3と図3に示すように、この意味フレームのスロットに“we”と“visit”と“Mito-Park”を埋め込めば、意味解析は終了である。結局、上に示した英文は表3と図3に示されるようにPSデータとMWデータのような形で、コンピュータ上で表現されることになる。

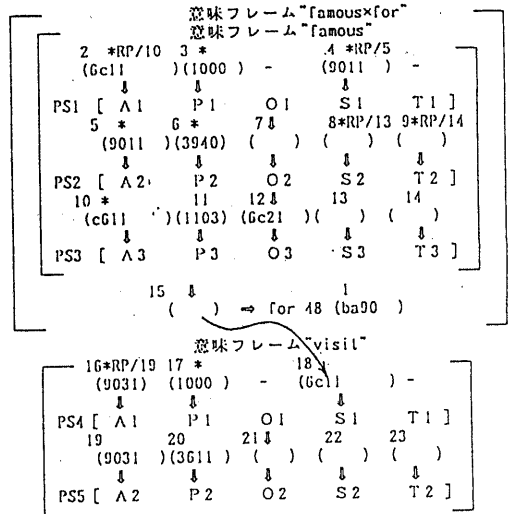


図3 (We visited Mito-Park that is famous for its plum-blossoms.) の構造文

6. まとめ

人間の頭の中に形成されてい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て、これによって自然言語の文法理論の構築を行っているが、人間はこの意味言語を用いて、質問応答、知識獲得、推論、学習、物語理解、翻訳などを行っていると考えられるので、この意味言語を用いて、このような人間の知的な機能をコンピュータ上で実現することを試みている。

本論文では、英文から得られた意味言語をマレー語と日本語に変換させることによって多言語間の同時機械翻訳が可能であることを明らかにした。本論文では次のようにプロセスで多言語の同時機械翻訳を実現した。

- (1) 英文をまず形態解析文法によって言語構造情報に分解して、言語構造情報テーブルであるWSテーブルを作成する。
- (2) それから作業テーブルであるMKテーブルを作成して、それを用いて、意味解析文法によって、意味言語のコンピュータ内での表現であるPSデータとMWデータに変換する。
- (3) このPSデータとMWデータから英語文とマレー語及び日本語文を作成する。

本論文で取り上げた英文はごく簡単な文であるが、複雑な文でも基本的には同じプロセスで実現できる。単語辞書、形態辞書、文生成辞書、形態解析文法、意味解析文法、自然文生成文法についてはほとんど触れなかったが、別の機会に各々について報告する予定である。

謝辞

本研究は長崎大学工学部機械システム工学科での卒業論文として協力された山本智之、M. ルトフィ、黒田修、黒田裕之らの諸君に感謝します。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昭教授(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。

参考文献

- (1) Chomsky, N., (1981): Lectures on Govern-

ment and Binding, Dordrecht: Foris.

- (2) Sell, P., (1985): Lecture 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) Fillmore, C., (1974): Toward a Modern Theory of Case and Other Articles, (田中春美, 舟越道雄訳, 1988, [格文法の原理], 三省堂.)
- (4) Ando, S., Goto, T. (1983): Current Status and Future of Intelligent Industrial Robots, IEEE Transaction on Industrial Electronics vol. IE-30 No.3 pp.291~299
- (5) 安藤司文, 1989; 認知プロセスと自然言語-自然言語の構造(1), 認知科学会, 第6回大会, 論文発表集, pp.128~129
- (6) 安藤司文, 1991; 意味言語の提案, 情報処理学会 自然言語研究会 情処研報 Vol.91, No.80, pp.73~80
- (7) 安藤司文, 1991; 自然言語から意味言語への変換, 情報処理学会第43回全国大会講演論文集(3), pp.3-137~13
- (8) 安藤司文, 1991; 意味言語による質問応答, 人工知能学会, ヒューマンインターフェイスと認知モデル研究会, 研究資料 SIG-HICG-9102, pp.43-52
- (9) 安藤司文, 1991; 自然言語に内在する知識獲得のメカニズム, 情報処理学会, データベース・システム研究会, 研究資料 情処研報 Vol.91, No.84, pp.1~10
- (10) 安藤司文, 1991; 意味言語による対話と学習, 認知科学会, 学習と対話研究分科会, SIGLAL, 91-2, 学習を対話, Vol.91, No.2, pp.1~10
- (11) 安藤司文, 1992; 自然言語に内在する論理構造, 情報処理学会 自然言語研究会 情処研報 Vol.92, No.2, pp.59~66
- (12) 安藤司文, 1992; 意味言語から見た日本文の構造-使役文と受け身文-, 国語学会平成4年度春季大会要旨, pp.63-68
- (13) 安藤司文, 1992; 意味言語から見た主語について, 言語学会春季全国大会, 資料