

知識表現FKR-Oとその機械翻訳への応用
KNOWLEDGE REPRESENTATION FKR-O AND ITS
APPLICATIONS TO MACHINE TRANSLATION

沢井 進 杉本 正勝 鞆詞 直哉
Sawai Sugimoto Ukai
Susumu Masakatsu Naoya

(富士通株式会社)

Fujitsu Limited

1. はじめに

日本語から英語というように自然言語を機械翻訳するには意味処理が必要なので¹⁾⁻⁴⁾、意味規則を効果的に記述する表現形式が求められている。本稿では、辞書/意味規則を表現する単純かつ強力な知識表現を一つ提案し、この知識表現を日英機械翻訳に適用した結果について報告する。

2. では知識表現FKR-Oについて述べ、3. ではFKR-Oで記述した辞書について述べる。4. ではFKR-Oの特長を述べ、5. ではFKR-Oを用いた日英機械翻訳について述べる。6. ではFKR-Oを用いて解決された点述べ考察を行う。

2. FKR-O

図-1はFKR-Oの枠組みである⁵⁾。

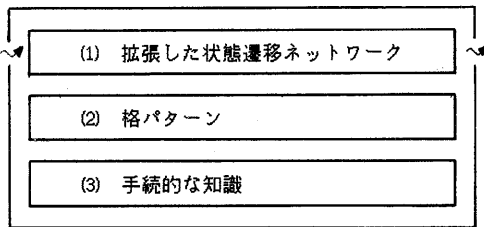


図-1 知識表現FKR-O

このようにFKR-Oは句型パターンを記述した格パターン(宣言的な知識)と、手続的な知識とを一対一対応に結びつけ、これをネットワーク化したものである。句型パターンはパターン照合処理を行うため、プロダクション・^(注)ルールの形式で書かれている⁶⁾。

(注) プロダクション・ルールと手続きを結び付ける方法は、Winogradや電子総合研究所の田中

穂積氏が既に着目している⁷⁾⁸⁾。 <1>

3. FKR-Oで記述した辞書

知識表現FKR-Oを用いる機械翻訳システムでは、意味規則をあらかじめ辞書の中に記述する。この辞書記述を効果的に行うため、「意味表現フレーム」と「制御フレーム」の2つのフレームを用意する。

3.1 意味表現フレーム

知識表現FKR-Oを用いる機械翻訳システムでは、意味規則をあらかじめ辞書の中に記述する。図-2はジャッケンドフによる動詞の意味表現で、「動作の主体(ACTOR)」と「動作(ACTION)」という主題関係を明らかにしている。具体的には、動詞「OPEN」が二つの名詞句を取ること、主語には「NP①」と「NP②」の二つの名詞句が成り得ること、NP②は「道具(INST)」であることを示している。ジャッケンドフは深層の動詞「CAUSE」を用い、主語が「対象(NP③)」を「開いている」状態にすることを示している⁹⁾⁻¹¹⁾。

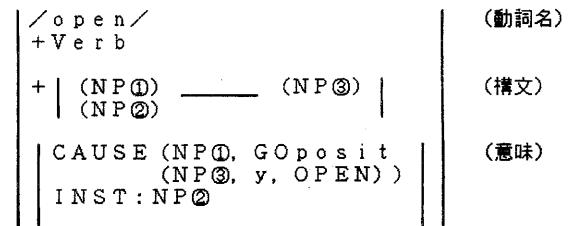


図-2 ジャッケンドフの意味表現

意味理論の最先端には「モンテギュー文法」等があり、最近研究されている。しかし、我々は図-2の「ジャッケンドフの意味表現」程度の意味表現があれば、意図する機械翻訳処理を行うのに十分であると考えている。

図-3はFKR-0で図-2の意味表現を記述したものである。これを意味表現フレームと呼ぶ。

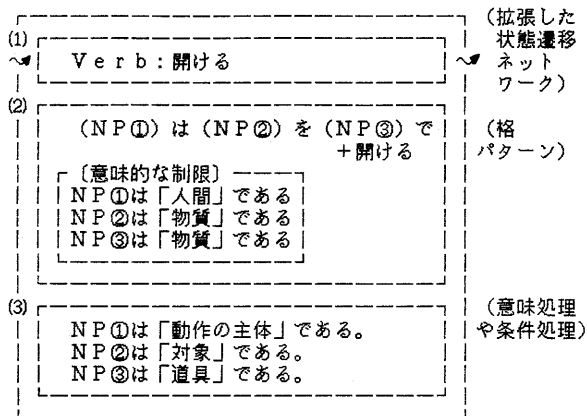


図-3 FKR-0による意味表現フレーム

(1)には状態遷移ネットワークの節(Node)名として「動詞名」があり、(2)には「動詞」と「名詞や名詞句」の相関関係、及び「名詞や名詞句」が守らなければならない意味的な制限がある。(3)は手続的な知識である。手続的な知識には、(2)の格パターン適用後に行う意味処理と条件処理がある。このような条件処理としては、条件を判定して別のフレームを呼び出すというような処理がある。

このようにFKR-0は図-2の「ジャッケンドフの意

味表現」を100%記述することができる。図-4はその一例である(2)。

Verb: 指定する				
	が	を	で	指定する
ルール1	NH (A)	NB (O)	NC (I)	VA
ルール2	NH (A)		NC (I)	VA

A (動作主)
O (対象)
I (道具)

(注) 「NH」は「人間」を, 「NB」は「抽象物」を, 「NC」は「物質」を意味する。「VA」は動詞「指定する」の原形を意味する。

図-4 意味表現フレームの一例

3. 2 制御フレーム

FKR-0を用いるシステムには、3. 1の意味表現フレームを制御する制御フレームがある。各フレーム間の通信は1つの制御レジスタで行っている。図-5のように、フレーム間の接続法には大別して3通りある。図-5の制御中枢は制御レジスタの内容でフレームを呼び出す役割を持っており、表面に現れない。次にどのフレームを呼び出すか明示していない場合は、制御中枢に戻る。

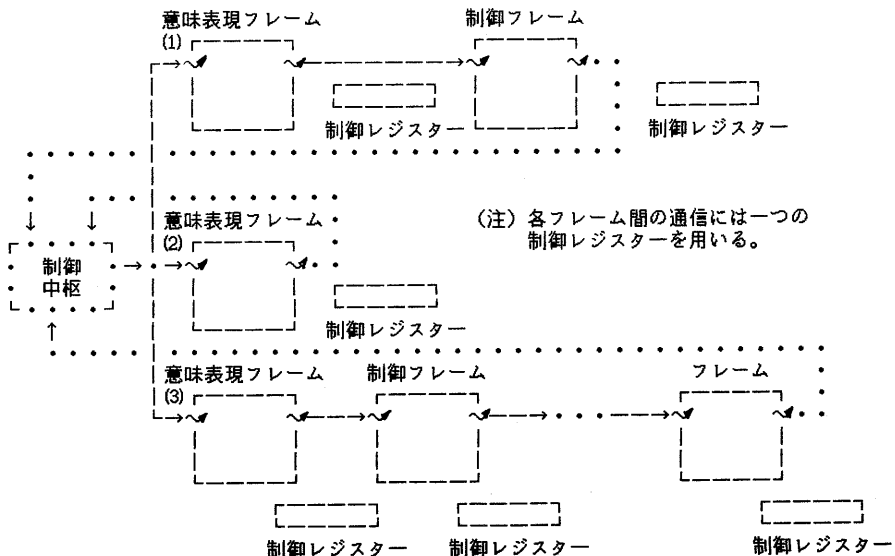


図-5 制御フレームと制御レジスタの役割

4. FKR-0の特長

知識表現は、知識工学で「専門家の知識」を表現し蓄積する道具として使用されている。代表的な知識表現にはプロダクション・システム（PS）、意味ネットワーク、フレーム、述語論理及び手続き（プログラム）がある¹³⁾¹⁴⁾¹⁵⁾。知識表現を機械翻訳に用いるには、「処理速度」、「単純さ」及び「相互作用性」を顧慮しなければならない。ここで、「単純さ」とは理解し易さと新しい知識の追加修正の容易さをいい、「相互作用性」とは制御情報を含めてヒューリスティック（経験的）な知識を容易に表現できることをいう¹⁴⁾。表-1はFKR-0と各種知識表現との違いを上記3点に関して比較したものである。

	処理速度	単純さ	相互作用性
PS	低速	単純	なし
意味ネットワーク	(注)	単純	なし
フレーム	(注)	単純	なし
述語論理	低速	単純	なし
手続き	高速	複雑	あり
FKR-0	高速	単純	あり

(注) システムの作りに依存する。

表-1 各種知識表現と比較

FKR-0は階層構造的に処理するので高速である。また、プロダクション・システムの単純さと、手続き特有の表現し易さを合せ持つ。

知識を如何に表現するかについて、MITのWinstonは「ある知識は手続き的である。たぶん考え方の集合は、プロダクションの集まりとしてとらえるのが最もよいであろう。その他の知識は叙述的で、簡単な表から複雑なフレームシステムにわたる広範囲な表現法のどこかに属す。たぶん直接に、各手続きを呼ぶほうがよいかもしれない。・・・」と述べている¹⁶⁾¹⁷⁾。

この点FKR-0はプロダクション・システムと関連する手続きとを直接結び付け、一つのフレームと考えており、強力な知識表現であると確信する。

5. FKR-0を用いた日英機械翻訳

5.1 機械翻訳システムFATRA

今回作成した機械翻訳システムを「FATRA」と呼ぶ。FATRAはFKR-0の内容を書換えながら徐々に成長できる機械翻訳システムで、「かな漢字交じり」の日本語を英文に変換する。

5.2 日英機械翻訳

FATRAによる機械翻訳は、日本語の入力、変換及び、英文の編集という3つのプロセスから成っている。当面、100%高品質完全自動翻訳するのは難しいので、英文の編集プロセスを持っている。さらに、FATRAには次の3つの機能がある。

- (1) 一文の翻訳中、訳文が出せなくなった時は途中までの翻訳結果を出す。
- (2) 英文を出力できない場合は原文を表示して、使用者に手助けしてもらう。
- (3) システムが辞書として使える。

FKR-0を用いる機械翻訳は、日本語の構文解析、意味（格）解析、言語の構造変換及び英語の合成を一体化させている点に特徴がある¹⁸⁾。図-6のように構文解析しながら、動詞に対する名詞の役割を調べている。動詞に対する名詞の役割を「格」といい、格関係を規則化した格パターンを参照し、意味を抽出する¹⁹⁾。一致する格パターンがあれば日本語の構造から英語の構造に変換するため、英語の意味合成、構文合成を行い合成した英文を出力する。

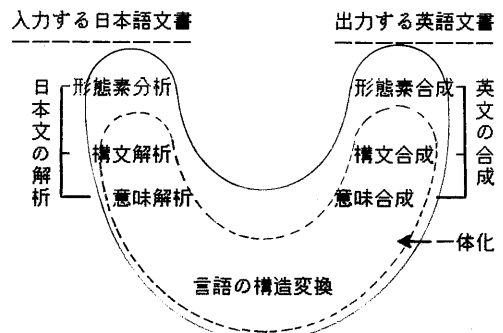


図-6 FATRAの全体フロー

これらの格パターンはFKR-Oの形式で辞書の中に書いてあり、条件に合う格パターンだけが英文合成に用いられる。図-6の点線部分はFKR-Oを用いて実現した部分である。

図-7の「かな漢字交じり」の日本語をFATRAに入力すると、FATRAは形態素分析を行い日本語を分ち書きする。その後、構文解析、意味解析、言語変換及び英文合成を行い英文を出力する。

(a) 形態素分析²⁰⁾

形態素分析は、前処理、熟語辞書引き、単語辞書引き、付属語解析及び固有名詞処理からなる。

・前処理

読点、「?」、「!」等で一文を切出す。同時に句点、カッコ等の特殊文字をパターン化する。

・熟語辞書引き

熟語辞書引きでは、慣用句、複合語、合成語並びにひらがな交じりの語句に属性と意味表示をつける。熟語辞書引きで引き当たらない場合は、漢字やカタカナの先頭に空白を挿入し、分ち書きする。

・単語辞書引き

単語辞書を引く、漢字やカタカナに始まる単語に属性と意味表示をつける。

・付属語解析

付属語解析では、熟語辞書や単語辞書を引き終わった後の文字列を調べ、付属語を抽出しパターン化する。付属語解析で引き当たらない場合は、かな辞書を引く。

・固有名詞処理

かな辞書にも無い語句は固有名詞のパターンを付ける。数字は、固有名詞として扱っている。

(b) 構文解析、意味解析、言語変換及び英語の合成

・構文解析

構文解析では、名詞の連結をFKR-Oの格パターンに従って処理したり、複文をFKR-Oの格パターンに従って単文の順序を並びかえたりする。

・表層格解析

分ち書きした文字列の名詞と助詞との関係がFKR

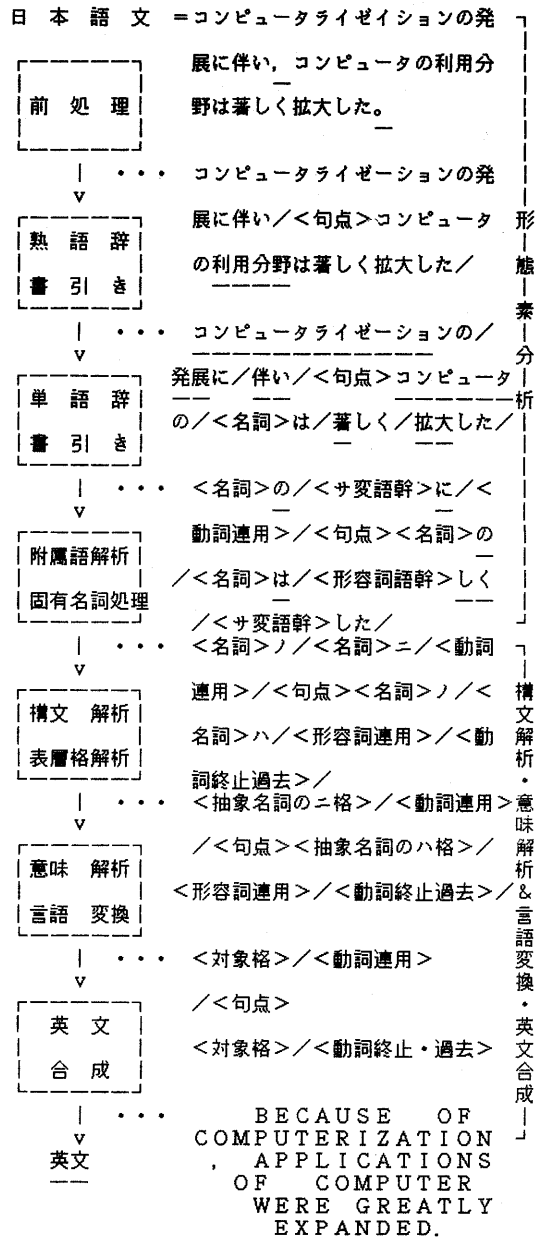


図-7 日英機械翻訳の一例

-Oの格パターンと一致するかどうか解析し、一致すれば表層格を生成する。

・意味解析

意味解析では、動詞の格支配を中心とするパターンとFKR-Oの格パターンとが一致するかどうか調べる。(パターンマッチング) このパターンマッチング

により、表層格から深層の意味を抽出する。抽出すべき意味には通常我々が5W1Hといっている「何時どこで誰が何をどうする」がある。FKR-Oで記述した辞書を用いることにより、時は「何時」なのか、場所は「何処」なのか、動作主とは「誰」なのかとか、対象は「何」なのかなどが抽出できるように成っている。

・言語変換

FKR-Oには格パターンが一致すると対応する英文のパターンが生成できるような日英対応表が手続き部にある。この対応表に従って、英文合成プログラムを呼び出す。

・英文合成

前置詞や定冠詞等、不足している語句を補いながら英文を合成する。複数形の処理もここで行う。

ただし、英文合成のうち形態素合成の部分はまだ作成していない。

6. まとめ

以上のことから、次の点が明らかになった。

(1) 高速処理

FKR-Oを用いればプロダクション・システム(P S)の欠点である認識部分のオーバーヘッドを減らすことができる。これはプロダクション・システムにフレームの階層構造の考えを導入し、制御フレームを用いることで実現した。表-2にFKR-Oを用いた機械翻訳と、プロダクション・システムだけを用いた機械翻訳とを比較する。

(2) 単純さ

既存のフレーム記述言語より知識表現モデルが単純である(2022)。

(3) 相互作用性

格パターンに「プログラムによる知識」を付加することで、格パターン間の矛盾が解決できる。このため新しい知識の追加修正がより容易にできる。

パターン 処理法	動詞がn個ある時の 機械翻訳の処理速度
FKR-O	$TS + \text{MAX}(T_i)$ ($i = 1 \text{ TO } n$)
PS	$TS + (T_1 + \dots + T_n)$

TS: 動詞以外の機械翻訳に要する時間

Ti: 動詞の機械翻訳に要する時間

表-2

(4) 知識の追加修正の容易さ

格パターンがフレーム形式でモジュール化でき、言語変換情報の追加や修正が簡単に行える。

(5) パターンマッチングの強力さ

一つのパターンマッチングの評価結果を利用して別のパターンマッチングを行うという強力なパターンマッチングも制御フレームを用いることにより行える。

FKR-Oのような枠組みを作成しておけば、徐々に翻訳分野の専門家が新たな知識を入力でき⁴ので、機械翻訳システム全体の能力を向上させ得ると考えている。

謝辞

いつも御指導いただいている開発事業部の佐藤事業部長、鈴木部長、村田部長、オフィス事業本部ソフトウェア事業部開発技術部の神田部長、古田課長に深謝いたします。

参考文献

- 1) Hutchins, W. J. : Progress in Documentation-Machine Translation and Machine-Aided Translation, Journal of Documentation, 34, 2 (Jun. 1978)
- 2) 長尾, 辻井ほか: 技術論文表題の英和自動翻訳の試み, 情報処理学会計算言語学研究会資料19-2 (昭54)
- 3) 石原, 田町: D-treeとそれに基づく英日機械翻訳のための言語解析について, 信学論(D), Vol. 57-D, No. 7, pp. 435-442 (1972)
- 4) 北川ほか編: 言語と情報-2, 講座情報社会科学, 第4巻, 学研
- 5) Minsky, M. : A Framework for Representing Knowledge, MIT AI Memo No. 306 (Jun. 1974)
- 6) 神田, 沢井, 横山: 自然言語処理用の辞書の構成法, 昭和54年度電子通信学会総合全国大会, 1206 (昭54)
- 7) Winograd, T. : Frame Representation & Understanding-Studies in Cognitive Science, (Bobrow, D. G. & Collins, A., eds.), pp. 185-210, Academic Press, New York (1975)
- 8) 田中穂積: 計算機による自然言語の意味処理に関する研究, 電子技術総合研究所研究報告, 第797号 (昭54)
- 9) Jackendoff, R. : Grammer as Evidence for Conceptual Structure, Linguistic Theory and Psychological Reality, MIT Press, pp. 201-228
- 10) 月刊言語: 特集・意味論入門, 12月, 1978
- 11) Natural Language Understanding, Stanford Heuristic Programming Project Memo HPP-79-21 (Jul. 1979)
- 12) 神田, 沢井: 機械補助翻訳を用いたドキュメント処理, 昭和54年度情報処理学会第20回全国大会, 3C-4 (昭54)
- 13) Bobrow, G. : A Panel on Knowledge Representation, Proceedings of the Sixth IJCAI, 1977, pp. 983-992
- 14) 田中(幸): 知識ベースとその応用, 情報処理, Vol. 21 No. 12, pp. 1231-1241 (1980)
- 15) 瀧 監訳: 人工知能の基礎-知識の表現と理解, 近代科学社 (1978)
- 16) Winston, P. H. : Artificial Intelligence, Adison Wesley (1977)
- 17) 長尾, 白井共訳: 人工知能, 培風館 (1980)
- 18) 神田, 杉本, 沢井: エンドユーザのための日本語によるプログラミング, 情報処理, Vol. 21, No. 3 (1980)
- 19) Fillmore, C. : The Case for Case, in Bach and Harms (Eds.) : Universals in Linguistic Theory, Holt, Rinehart, and Winston (1968)
- 20) 絹川: 情報検索のための日本語解析, 情報処理, Vol. 20, No. 10 (昭54)
- 21) Robert, R. B. and Goldstein, I. P. : The FRL Manual, MIT AI MEMO 409 (Sep. 1977)
- 22) Bobrow, D. G. and Winograd, T. : An Overview of KRL, A Knowledge Representation Language, J. Cognitive Sci., Vol. 1, No. 1, pp. 3-46 (1977)