

格文法による仮名漢字変換の多義解消†

大 島 義 光** 阿 部 正 博**
湯 浦 克 彦** 武 市 宣 之**

べた書き文の仮名漢字変換における問題点——同音語による多義と形態素分割による多義——を、格文法をベースとする構文意味解析を用いて統一的に解消する方法について述べる。格文法は意味情報を用いて同音語の選択を行うことができ、仮名漢字変換への応用には有効であるが、そのためには分解能に十分配慮する必要がある。ここでは、国立国語研究所の分類語彙表に11種のマクロ意味コードを加えた意味体系と、意味分解能、一文一格の原理、変形処理における格の一貫性を考慮して定義した22種の格による格文法モデルを採用して、これに対しての処理は、形態素解析によって網羅的に生成した形態素候補群を入力とし、そこから格文法により分解可能な形態素多義を含むパス（文候補）を一つずつ取り出して構文意味解析を適用し、日本語文としての尤もらしさを評価することにより、形態素分割多義の判定を行う。また、各パス内では単文要素に対する格フレーム照合を中心とし、係り受けのチェックおよび意味による同音語の選択を行うが、飛びこし係りの処理、連体修飾用言の被修飾名詞に対する同格連体修飾の処理などにより正確な処理をはかっている。実験システムを構成し評価実験を行った結果、総合の仮名漢字変換率が約95%、このうち構文意味解析の寄与は約1%であった。

1. ま え が き

最近のOAの進展に伴い日本語ワードプロセッサやパーソナルコンピュータが急速に普及しつつあるが、これらの機器では日本語の入力手段として仮名漢字変換が広く用いられている。現在実用に供されているものの多くは文節分かち書きによる仮名漢字変換^{1)~4)}であるが、これにはいくつかの問題点がある。

まず、日本語には分かち書きの習慣がないため、文節ごとに区切って入力するのはユーザにとって不自然であり、手数も増えるという問題がある。また、文節の概念には細部で不明瞭な部分(補助動詞や形式名詞、複合語などの取り扱い)があるため、システムのルールとユーザの予想にずれを生じる点も問題である。これを解決するには、区切り入力の不要なべた書き入力の仮名漢字変換を実現する必要がある。

次に、仮名漢字変換の変換結果には多かれ少なかれ誤りが含まれるが、これを修正する手間も無視できない。さらに、べた書き入力では、分かち書き入力に比べて入力文の曖昧さが多くなるので、変換誤りも増大する。操作性の向上や入力工数の削減のために変換率の向上が望まれる。

前者のべた書き入力の問題については、近年いくつかの解決法が提案されている^{5),6)}。後者については従

来頻度情報ないし頻度の学習機能を用いるのが一般的であるが、これには限界がある。一層の精度向上を望むためには、構文や意味、さらには文脈情報など、より高度の情報を用いた処理が必要である。

これまで構文や意味の情報を利用する方式として、すでに用言と体言の格関係や関連語情報を利用するのが報告されている⁷⁾。これは構文意味情報を同音語の選択に利用したものである。

我々は、べた書き文の仮名漢字変換システムを構築し、その中で構文意味解析を同音語の選択だけでなく形態素分割の多義解消にも適用することを試みた。実験の結果、その有効性が確認されたので、以下その内容について報告する。

まず第2章で仮名漢字変換における多義の内容について概観し、その解消方法について考察する。次に、第3章で我々が用いた言語モデル、具体的には格文法モデルの内容について述べ、第4章ではこれに基づく構文意味解析を解説し、最後に実験結果について報告する。

2. 仮名漢字変換における多義とその解消方法

文節分かち書きの仮名漢字変換では同音語や品詞・活用などの多義が変換誤りの要因となっていたが、べた書き文の仮名漢字変換では、さらに単語の区切りの多義も変換誤りの要因となる。これらの多義の例を表1に示す。

このような仮名漢字変換の多義を、構文意味解析を

† A Disambiguation Method in Kana-to-Kanji Conversion Using Case Frame Grammar by YOSHIMITSU OSHIMA, MASAHIRO ABE, KATSUHIKO YUURA and NOBUYUKI TAKEICHI (The 8th Department, Central Research Laboratory, Hitachi Ltd.).
** (株)日立製作所中央研究所第8部

表 1 仮名漢字変換における多義の例
Table 1 Various ambiguities in kana-to-kanji conversion.

分類	多義の種類	例	*1
同音同品詞	漢字/仮名表記	移す/うつす	
	送り仮名	行なう/行う	
	同音同義語	追究/追窮	
	同音類義語	製作/制作, 定跡/定石	
	同音異義語	構成/厚生, 買う/飼う	○
品詞活用多義	名詞/形容動詞	政党に/正当に	○
	動詞(活用多義)		
	①音便による同形化	読んだ(マ行)/呼んだ(バ行)	○
	②一段動詞/可能動詞	変える/買える	○
	名詞/動詞連用形	又借り/またがり(跨り)	○
形態素分割多義	副詞/動詞	挙げて/上げて	○
	接辞/助詞	化/か, 派/は, 氏/し	○
	複合語内境界	中心・性・測地 生息・地	
形態素分割多義	後続文節の先頭を助詞と誤認。または、その逆。	群れからは慣れる 離れる 出席し定見を述べて意見	○
	複雑なもの	副夢想/含む層 計画を楯状教を把握し立て状況	

*1: ○印は本稿で多義解消の対象としたものを示す。

用いて精度良く解消していくには、まず、変換対象である仮名文から正解となる可能性のある形態素を見逃すことなく抽出する必要がある。次に、この形態素の候補に対して、適切な構文意味情報を適切な手順で適用することにより、正しい変換結果を得るとというのが基本的な問題の解決方法である。そこで、我々は次のようなアプローチで問題解決をはかることを試みた。

まず、形態素解析により、入力された書き文から有意な形態素候補を網羅的に取り出す。このとき、正解候補は余すことなく取り出すが、無用な候補を際限なく出力するのは、以後の構文意味解析でノイズとして働く可能性があるため抑制する必要がある。形態素解析の処理の詳細はここでは割愛するが、このために複合語の接続尤度に関する知識などを利用している⁹⁾。

次に、形態素解析で生成した形態素候補群から文候補を一つ一つとり出し、構文意味情報を適用して日本語文としての尤もらしさの検定を行い、適切に形態素分割された正解文候補を得る。また構文意味解析の過

程で意味の情報を用いて同音語の選択を行う。

以上の処理を実現するために、次の情報を用いる。

(1) 構文的には、日本語の文節間に働く係り受け規則と、それに関連する若干のヒューリスティック規則。

(2) 意味的には、用言と体言の間の意味的関係を規定する格関係、すなわち格文法^{9),10)}。

格文法は、これを利用することにより意味を用いて同音語の選択を行うことができ、仮名漢字変換への応用には有効である。そのためには格文法処理に必要な格フレーム辞書作成にあたって、意味による同音語分解能が十分高くなるように配慮する必要がある。

3. 格文法モデルおよび格フレーム辞書

格文法を利用するのに必要な、意味体系、用言の格の種類など、格文法モデルの仕様について述べる。

3.1 意味体系

仮名漢字変換用の意味体系には詳細かつ汎用的な意味体系が必要であるが、これに合致する既存のものとして、国立国語研究所発行の分類語彙表¹⁶⁾、角川類語新辞典¹⁷⁾などがある。このうち分類語彙表は、同義語・類義語の意味範囲の区分が比較的厳密に作られており、仮名漢字変換への応用に適すると考え、これを我々の意味体系のベースとすることにした。しかし、階層性、体系性の面で若干不満のところがあつたので最下位の分類を中心に使用し、不足の部分を後述のマクロ意味コードで補っている。自立語辞書中の単語の意味コードには、この最下位の分類を付与している。また、格フレームの各格スロットに付与する意味コードにも最下位の分類を用い、これを複数並記することを原則にしたが、必要な場合は上位コードも付与可能としている。

マクロ意味コードは、格フレームの意味記述に頻出する一群の意味コードをまとめたものである。「人」「もの」「場所」など一般的概念を指す意味コード群で、分類語彙表の体系の中で関連するものをまとめ、一つの意味コードとして扱えるようにした。

このマクロ意味コードはまた、指示代名詞(「これ」「それ」「ここ」「そこ」など)や形式名詞(「こと」「もの」など)などの、表す意味が広い格フレーム内の意味記述では対応しきれない指示語類にも付与し、意味処理の精度向上に寄与させている。

マクロ意味コードの一覧を表2に示す。

表 2 マクロ意味コード
Table 2 Macro meaning codes.

格フレームに記入するもの	人間 (5. 11), 組織 (5. 12) 有意志体 (5. 1=5. 11 5. 12), 時 (5. 2), 場所 (5. 3), 行為 (5. 4), 物 (5. 5), 感情 (5. 6), 現象 (5. 7)
指 示 語	形式名詞 「こと」 (5. 9=5. 4 5. 6 5. 7), 「もの」 (5. 5)
	準体助詞 「の」 (5. 8=5. 5 5. 9)
	代名詞 「これ」「それ」 etc. (5. 8), 「ここ」 etc. (5. 3)

* 先頭の '5' はマクロ意味コードのマーク。

3.2 格の定義

3.2.1 深層格

これまでに、種々の格体系が提案されているが¹¹⁾⁻¹⁵⁾、本システムでは、意味処理による分解能を重視して、格の設定は深層格(意味のレベル)で行うこととし、格助詞(表層格)をこれに付記する形式とした。

一文一格の原理*などを考慮して、以下の22種の格を定義した。

主格, 対象格, 場所格, 起点格, 終点格, 時間格, 開始時間格, 終了時間格, 手段用具格, 材料格, 相手格, 比較の基準格, 資格格, 引用格, 提題格, 補格, 程度数量格, 結果目的格, 原因理由格, 条件場合格, 使役主格, 受身主格。

3.2.2 表層格

表層格は、格助詞および「に関して」「によって」などの格助詞相当句、時間表現・数量表現などに現れる助詞を伴わない名詞文節に対応するための空連糸の助詞、また、「象は鼻が長い」などの表現に対処するための提題を表す係助詞「は」、などを用いている。

3.2.3 格の必須性

任意の用言について、意味的に不可欠な格を必須格、そうでない格を自由格として区別した。また、不正な係り受け処理の防止をはかる意味から、個々の用言の格にはなりえない格を不要格として格フレーム中に明示している。

3.3 変形情報

動詞・助動詞に「れる」「られる」がついて受身文になると、用言と格の関係が変化する。これは表層格である助詞の交替という形で起こる。

(例) Aが Bを Xする

→ Bが Aに(よって) Xされる

* A: 主格, B: 対象格, X: 用言

受身に関するこのような格フレームの変化を、18通りのパターンに分類し、この変形パターンのコードをそれぞれの格フレームに付与している。

「れる」「られる」はまた、尊敬・可能および自発の意味を表す機能がある。これらはそれぞれ受身とは別の構文形態をとる。尊敬の場合は、基本形をそのまま用いれば良いが、可能の場合は、

「魚を食べる」→「魚が食べられる」

のように助詞の交替が起こる。また、自発の場合は、

「私は～と思う」→「私には～と思われる」

のような変形が起こる。

ただし、すべての動詞が尊敬・可能・自発および受身の用法を完備しているわけではなく、動詞の意味・用法ないし活用の種類によっては、いくつかの用法を欠くものがある。これを明示するために、それぞれその有無を示すフラッグを格フレーム中に設置した。(受身の有無は変形パターン・コードで指定する。)

「せる」「させる」がついて使役文になる場合も受身と同様にして取り扱う。使役に関しては10種の変形パターンを用意している。

3.4 格フレーム辞書

以上の基本仕様に従い、格フレーム辞書を作成した。格フレームを作成する用言の選定は、主として国立国語研究所の新聞語彙調査資料¹⁸⁾をもとに、そこから新聞特有の語を除き、これに電気・電子関係の技術用語を若干数追加して作成した。うちわけは、和語動詞1823語、サ変動詞1924語、形容詞254語、形容動詞587語である。

格フレーム辞書の内容例を図1に示す。

4. 構文意味解析

次に、構文意味解析の処理について説明する。

4.1 形態素多義の表現

図2に形態素解析で生成した形態素候補群の例を示す。形態素解析の処理の効率化、および多義の効率的保持のためにネットワーク構造を採用した。ネットワークの各々の枝は、形態素の分割位置ないし品詞・活用別に生成する。同一品詞、同一活用の語が複数あるときは、同じ枝の下に同音語としてつながれる。

このネットワークからパス(文候補)を一つずつ取り出し、それに対して構文意味解析を適用する。

* 一単文中に同一の格はただか一度しか現れないという経験則。

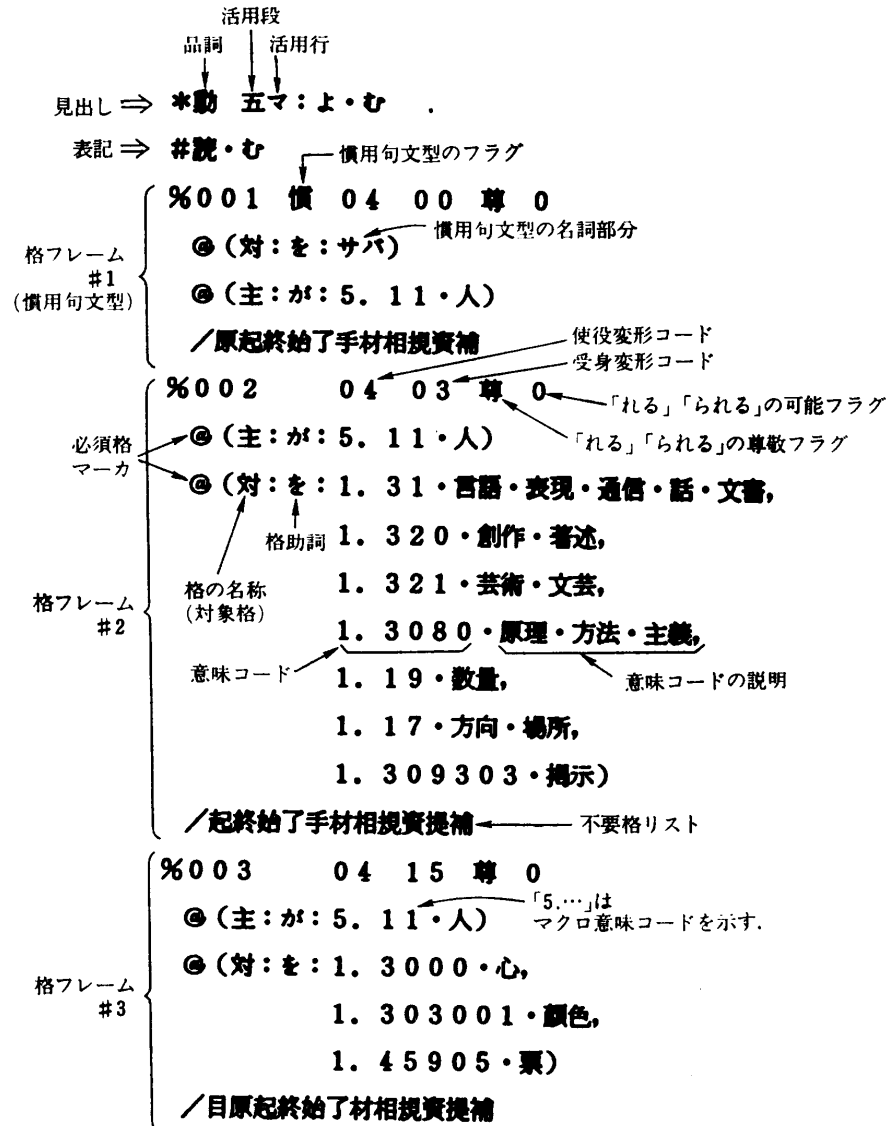


図1 格フレームの記述例
Fig. 1 An example of case frame.

4.2 パス (文候補) の抽出

表1に示したように、上記形態素ネットワーク上に存在する多義の種類は広範に渡っている。一般に、このようなネットワークから可能なパスをすべて切り出すと、その数は膨大なものになる。これらすべてのパスに構文意味解析を適用して尤度を判定するのは、処理時間の点で問題であるし、構文意味解析の負担も大きすぎる。そこで、あらかじめヒューリスティックな手法を使って取り扱う多義の範囲を限定し、構文意味解析で有効な判定ができるような多義に絞る。

まず、パスを構成する自立語数、局所的構文尤度、自立語の頻度の和などの情報を利用して、パスの候補

を一つに絞る⁸⁾。次にこのパスの周りで表1中の○印を付けた多義を探し、それを含むパスを対立パスとして生成する。第一のステップは最も可能性の高い候補を抽出すること、第二のステップはそれによって落とされる多義のうち、格文法処理により回復可能なものを復活するステップとなっている。このようにして抽出されたパスに対して各々構文意味解析を適用し、その処理結果から最適パスを選択する (図3参照)。

格文法処理による多義処理の対象としたもの (表1中の○印) は、同音意義語、品詞・活用による多義、および形態素分割多義の一部である。形態素分割による多義には種々雑多なものがあり、すべてを多義処理

の対象とすると却って変換誤りを増大させる危険性がある。そこで、「は」「も」「か」など単語の切目の多義になりやすい短い助詞を多義処理の対象とした。

4.3 文節構造の抽出

次節の格フレーム照合処理に先立ち、照合単位となる文節内のデータの整理を行う。

4.3.1 文節品詞属性の決定

文節の品詞属性は、通常は自立語の品詞によって決定されるが、動詞の連用形(例:「書き換え」:動詞→名詞)、名詞+接辞「化」(例:「製品・化する」:名詞→動詞)、名詞+接辞「的」(例:「経験・的な」:名詞→形容動詞)などは品詞を認定しなおす必要がある。

4.3.2 係り型

後続文節への係り型(連用、連体、並立)は文節末尾の形態素の形からおおむね判定できる。ただし、助詞の「と」については連用と並立の多義があり、文節内処理では決定できないので、後の格フレーム照合処理の中で判定する。「の」は通常連体修飾を表すが、連体形用言の前では連用修飾とする。

4.3.3 文節代表助詞の認定

格フレーム照合時、名詞文節に付随する助詞を格のキーとして用いる。そこで文節の付属語中から代表助詞を一つ決定する。まず、格助詞を優先し、格助詞がなければ最後に現れた助詞(副助詞または係り助詞)を取る。時間表現または数量表現における助詞の付かない名詞文節の場合は、空連糸の格助詞を仮定する。

4.4 格文法による係り受けおよび意味解析

解析は、用言と連用名詞文節からなる単文に対する格文法処理が基本となる。

文頭側からパス(文候補)を走査し、用言文節以外の文節を順にスタックへ積み。用言文節が現れたところでその用言を中心とする格文法処理を行う(図4)。

4.4.1 用言と連用名詞文節の格文法処理

まず、用言の格フレームを辞書から読み出す。このとき用言文節の付属語部分に格フレームの変形を引き起こす付属語(助動詞、補助動詞)があるときは、変形パターンを用いて格フレームに変形を施し、変形格フレームを作成する。また、不要格情報をもとに不要格に挙げられていない自由格をプログラムで補充す

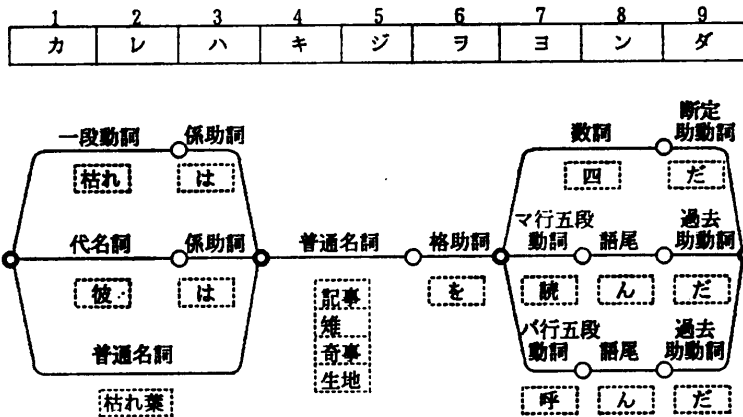


図2 形態素ネットワークの例
Fig. 2 Example of morpheme network.

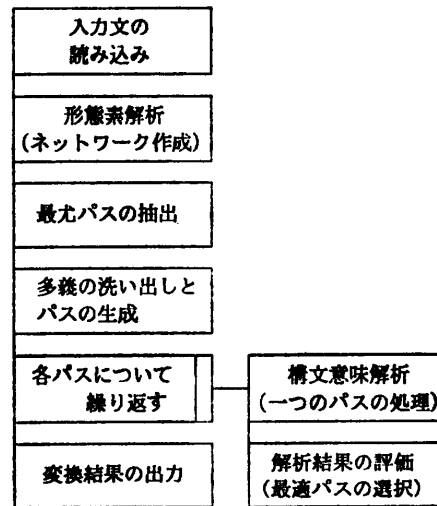


図3 仮名漢字変換制御手順
Fig. 3 Main control flow of kana-to-kanji conversion (Depicted with PAD).

る。

次にこの格フレームとスタック内の名詞文節との照合を行う。照合はスタックの先頭すなわち用言に最も近い文節から始め、文頭へ逆上る方向で一文節ずつ処理を進める。これにより係り受けの方向性と係り受け非交叉の原理を保持しつつ、解析を進めることができる。文節と格の照合は、まず文節の助詞をキーに格の引き当てを行い、次に意味コードの照合を行う。文節の代表助詞が格助詞(または助詞相当句)のときはそれをキーとするが、副助詞・係り助詞のときは適宜「が」または「を」に読み換えて照合する。

意味照合の過程で意味情報をもとに同音語の選択を行う。また連用名詞文節と並立関係にある名詞文節に

ついても、意味コードにより同音語の選択をする。

なお、文節と格の照合において、既に他の文節へ引き当て済みの格は用いず、未照合の格を使用する（一文一格の原理の適用）。

4.4.2 連体形用言の処理

用言が連体形の場合は、その被修飾名詞に対して格フレーム照合を試みる。ただしこのときは助詞の情報は用いず、意味コードのみにより照合を行う。

(例) 私が買った本 (対象格)

なお、一部の名詞は、連体修飾用言に対し、格とは別の関係をとることがある。これらは「計画」「発明」「状況」「経験」など、広く言ってコト的な意味を持つ名詞で、直前の用言が構成する単文全体を指している（同格連体名詞¹⁹⁾）。このような名詞の処理も正しく行えるよう、品詞の下位属性で識別可能にしている。

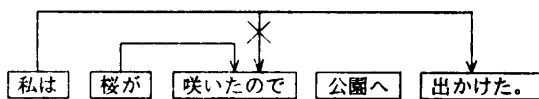
(例) 新製品を開発する計画

4.4.3 飛びこし係りの処理

通常、連用名詞文節は最も近い用言文節に係ることが多いが、場合により直後の用言を越えてより後方の用言に係るものがある。このようなものを処理するために、次に示す二つの機構を用意した。

(i) 格フレーム照合の過程で係り先が決まらずにスタック中に残った文節をより後方の用言の処理にまわす。

(例)



(ii) 「～は」という表現は文の主題などを提供し、通常連体形の用言に係ることはない。連体形用言の処理時スタック中に「～は」という文節が現れたときは、そこで格フレーム照合を取り止め、スタック中に残る文節を次の用言の処理にまわす。

(例)

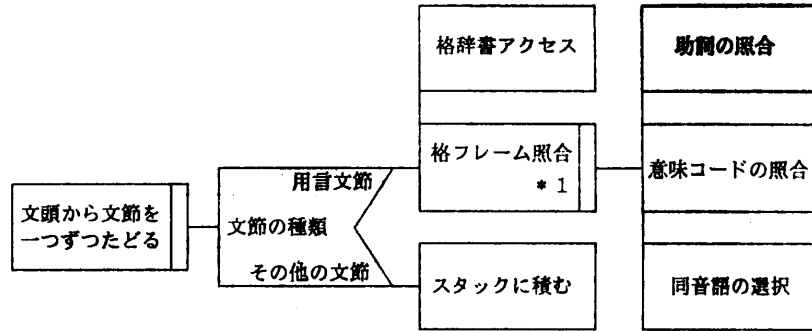
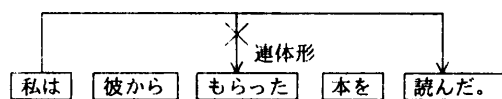


図 4 一つのパスの構文意味解析
Fig. 4 Flow diagram of syntactic and semantic analysis of a path (Depicted with PAD).

*1 用言の同音語、各用言の異なる格フレーム、格と文節の組合せ、など、すべての多義について照合を行う。

4.5 照合結果の評価

4.5.1 パス内での評価

パス内では、各用言ごとに格フレーム照合評価を行う。このとき、用言の同音語、各用言の異なる格フレーム、格と文節の組合せ、などすべての多義について照合を行い、その中から最適なものを選択する。

評価の手順は、まず、必須格の多いものを優先し、次に総引き当て格数を見る。それでも同等のときは、引き当てた名詞文節と用言の単語頻度の総和の大きなものとする。

4.5.2 パスの評価

パスごとに行った構文意味解析の結果を用いて、その中から最適なものを選択する。

まず、格フレーム照合結果の良否についてチェックを行う。具体的には、各文節の係りおよび意味照合の結果をチェックし、解析を失敗した文節を含まないパス（完全パス）を優先する。

完全パスが二つ以上ある場合は、さらに各パスに含まれる必須格数をカウントし、これが多いほうが文として尤もらしいと判断し、必須格の多いほうを優先させる。これもまた同等のときは、各パスの全要素の頻度を比較し、大きいほうを選択する。

5. 実 験

5.1 実験システム

実験システムの構成を図5に示す。使用した計算機は HITAC M-200 H, プログラム言語は PL/I である。自立語辞書ならびに格フレーム辞書には VSAM (Virtual Sequential Access Method) ファイルを利用した。

自立語辞書は固有名詞約3万5千語を含む約16万語の自立語からなる。それぞれ、読み、表記、品詞、活用型、使用頻度値、分類語彙表の意味コード、などを付与してある。格フレーム辞書は約4600語の用言からなり、用言の語幹の読みと自立語辞書の同音語番号をインデクスとして検索することができる。

5.2 解析例

図6に解析の一例を示す。4章の手順に従って解析が行われ、格ラベルを付した係り受け関係と同音語の選択結果が得られる。これは一つのパスに対する処理例であるが、形態素ネットワークから抽出されたパスが複数あるときは、各々のパスについて同様な処理が行われ、その中から最適なものが選ばれる。

5.3 実験結果

技術文書・事務文書・科学雑誌記事などからなる総計10150字(仮名文字数)のテキストデータを用いて評価実験を行った。テキストに含まれる文の総数は197文、文節数は1987文節、うち格文法処理の対象となる文節数は1245文節(約63%)、また単文数は505文である。

実験は形態素解析単独の場合と構文意味解析を適用した場合の2通り行い、それぞれ仮名漢字変換率を算出した。結果を表3に示す。形態素解析単独の場合の値は、4.2節で最適パスを一つ取り出したときまでの結果である。

処理時間は1文節当たり平均91.6msであり、このうち構文意味解析付加による処理時間の増加は16.4ms(22%増)であった。

処理結果について、述語部と名詞部を比較すると、述語部は複数の情報(名詞)から支持されて選択されるため、正解となる比率が高くなっている。

変換誤りの原因については、各種言語データの不備、アルゴリズムの不備による部分もあるが、より本質的には次の諸点上げることができる。

(1) 構文意味解析は正しく処理しているにもかかわらず、該当候補が複数あり、正解候補が選択されなかった(1.0%)。

(例) 「{耕地, 高地, 高知}に住む」
(注: アンダーライン部が正解)

これを解決するには現在の単文を中心とする格文法処理では不十分で、文単位ないし複数文にわたる文脈

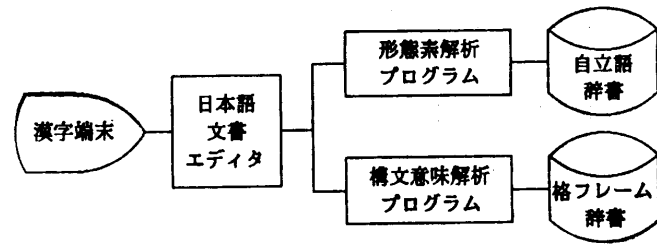


図5 実験システムの構成
Fig. 5 Organization of experimental system.

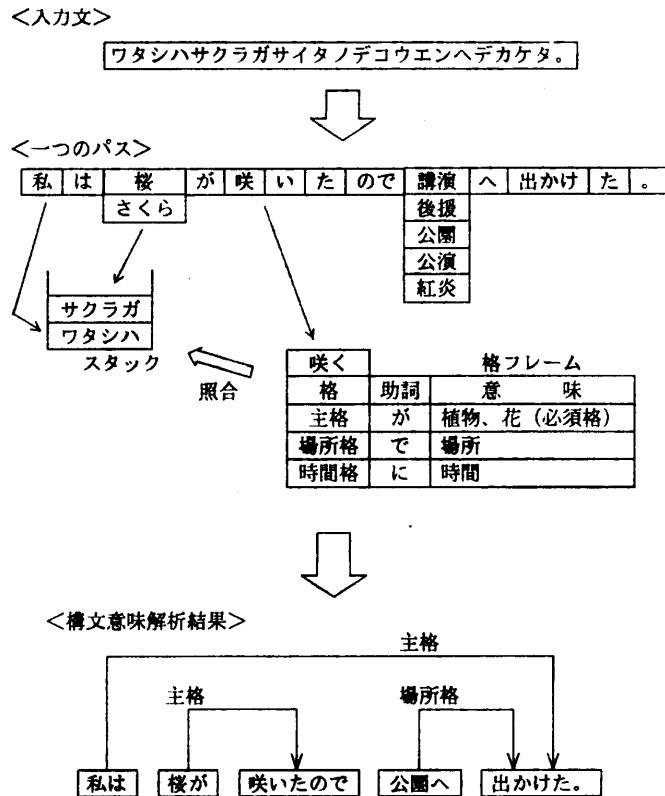


図6 構文意味解析の一例
Fig. 6 An example of syntactic and semantic analysis.

表3 実験結果
Table 3 Results of experiment.

	構文意味解析を適用した場合	形態素解析単独の場合
仮名漢字変換率*	94.9	93.9

* 形態素単位に正誤を判定したときの、カナ文字換算による正解率。

処理が必要である。

(2) 現在、意味処理の対象としていない副詞文節や連体修飾型の名詞文節、ならびに複合語の処理不足によるもの(1.2%)。

ゴリラの社会には創造以上^{想像}に複雑で厳格な起立^{規律}がある。

マウンテン・ゴリラは中央アフリカの耕地^{高地}に住むゴリラの亜種で、低温環境に耐えるためにからだが大きく体毛が長い。中心性測地バルンガ火山地帯はザイール、ウガンダ、ルワンダにまたがり、3国の国立公園に指定され保護区となっている。ここにゴリラの生体^{生態}を研究するカリンケ・リサーチ・センターが20年前に解説され、現在ダイアン・フォンシー助詞が自然な^{開設}環境のなかでの観察を続けている。

このゴリラ保護区には、マウンテン・ゴリラのほか、声部^{西部}低地に住むものと頭部^{東部}低地に住むものを合わせて三つの亜種が認められているが、それらの形態があまり大きく違っていないことから考えると、比較的近年に分散したものらしい。

人間による干渉や自然障壁のために、これらのゴリラ達はさらに小さな孤立した群れに分かれて森の凹みに住んだ。群れの大部分は橋の侵食や密漁等の厳しい圧迫に直面して、ますます細かく^{土地}分割^{密猟}され、体格も小型化していった。このような小さな群れの2, 3のものについては、固体数がある程度正確に測めているが、声部^{西部}と頭部^{東部}中央アフリカの固体数^{個体}については、その総数^{個体}がはっきりしていない。

図7 変換例

Fig. 7 An example of kana-to-kanji conversion.

*1 ゴリラの社会: NEWTON, Vol. 2, No. 7 より抜粋.

*2 アンダーラインは誤変換部を示す.

解決には処理範囲の拡大が必要.

(3) 専門語, 固有名詞, 略語など対象分野の知識を動員しないと解決できないもの, あるいは対象とする文書の書式の知識を必要とするもの(例: 章・節の見出し)(1.3%).

この解決は専門語辞書・個人辞書および特殊な書式処理による.

図7に, 変換結果の一例を示す.

6. むすび

仮名漢字変換の精度向上をはかるためには, 種々の制約条件(constraint)ないしはヒューリスティックな

手法を駆使して, 多義の範囲を限定する必要がある. ここでは格文法をもとにした構文意味解析と, パスの適用手順および結果の評価にヒューリスティクスを導入して, 約1%の仮名漢字変換率の向上効果を得ることができた.

今後は, 格フレーム辞書など言語データの充実と精度向上, ならびに「の」による連体修飾や複合語などへの処理範囲の拡大, さらに文脈情報などを利用した, より精密な処理などを導入して, 一層の精度向上をはかっていきたい.

謝辞 終わりに, 本研究を進めるにあたり, ご討論いただいた京都大学長尾真教授, 本研究の機会を与えてくださった(株)日立製作所中央研究所堀越彌企画室長(研究当時第八部長), 格フレーム辞書の開発にご協力いただいた(財)計量計画研究所木村睦子主任研究員, 構文意味解析アルゴリズムおよび格フレーム辞書の仕様の検討にご協力いただいた(株)日立製作所マイクロエレクトロニクス機器開発研究所中島晃氏(研究当時中央研究所第八部), 本研究のプログラム開発にご協力いただいた(株)日立超LSI横森久美子氏に深く感謝いたします.

参 考 文 献

- 1) 栗原, 黒崎: 仮名文の漢字混じり文への変換について, 九州大学工学集報, Vol. 39, No. 4, pp. 659-664 (1967).
- 2) 相澤, 江原: 計算機によるカナ漢字変換, NHK 技術研究, Vol. 25, No. 5, pp. 261-298 (1973).
- 3) 松下, 山崎, 佐藤: 漢字かな混じり文変換システム, 情報処理, Vol. 15, No. 1, pp. 2-9 (1974).
- 4) 河田, 天野, 武田, 森: ミニコンピュータを用いたカナ漢字変換システム, 電子通信学会技報, PRL 76-47, pp. 25-34 (1976).
- 5) 牧野, 木澤: べた書き文の分かち書きと仮名漢字変換—一文節最長一致法による分かち書き, 情報処理学会論文誌, Vol. 20, No. 4, pp. 337-345 (1979).
- 6) 吉村, 日高, 吉田: 文節数最小法を用いたべた書き日本語文の形態素解析, 情報処理学会論文誌, Vol. 24, No. 1, pp. 40-46 (1983).
- 7) 牧野, 木澤: べた書き文の仮名漢字変換システムとその同音語処理, 情報処理学会論文誌, Vol. 22, No. 1, pp. 59-67 (1981).
- 8) 武市, 阿部, 大島, 湯浦: 構文意味解析を適用したべた書き文仮名漢字変換システムの開発, 他, 情報処理学会第28回全国大会講演論文集, 4M-5-

4M-8 (1984).

- 9) Fillmore, C. J.: 格文法の原理, 田中, 船城(訳), 三省堂, 東京(1975).
- 10) 井上和子: 変形文法と日本語(上)(下), 大修館書店, 東京(1976).
- 11) 石綿: 日本語の生成語彙論的記述と言語処理への応用, 国立国語研究所報告 54, pp. 152-198 (1974).
- 12) 絹川, 木村: 日本語文構造解析による自動インデクシング方式, 情報処理学会論文誌, Vol. 21, No. 3, pp. 200-207 (1980).
- 13) 坂本: 格構造を中心とした用言と付属語辞書, 情報処理学会自然言語処理研究会資料, 38-8 (1983).
- 14) 島津, 内藤, 野村: 格構造モデルに基づいた日本語文の分析と解析, 情報処理学会自然言語処理研究会資料, 29-1 (1982).
- 15) 池田: 一般化された格構造による意味表現を用いた日本語文の構文解析法について, 電子通信学会論文誌, Vol. J60-D, No. 10, pp. 814-821 (1977).
- 16) 国立国語研究所: 分類語彙表, 秀英出版, 東京(1964).
- 17) 大野, 浜西: 類語新辞典, 角川書店, 東京(1981).
- 18) 国立国語研究所: 電子計算機による新聞の語彙調査(II), 国立国語研究所報告 38, 秀英出版, 東京(1971).
- 19) 奥津: 生成日本文法論, 大修館書店, 東京(1974).
(昭和60年9月30日受付)
(昭和61年5月15日採録)



大島 義光 (正会員)

昭和22年生。昭和46年東京大学工学部計数工学科卒業。昭和48年同大学院計数工学専門課程修士課程修了。同年(株)日立製作所入社。以来、同社中央研究所にて、漢字ディスプレイ、タッチパネル、リレー回路図と論理式の相互変換アルゴリズム、日本語ワードプロセッサ、仮名漢字変換などの研究開発に従事。電子通信学会会員。



阿部 正博 (正会員)

昭和23年生。昭和46年東京大学教養学部基礎科学科卒業。同年、(株)日立製作所中央研究所入社。以来、計算機システムの性能評価、計算機のユーザビリティ、自然言語処理、知識処理などの研究に従事。昭和54年8月より1年間、英国 Cambridge 大学 Computer Science 学科に留学, Diploma。現在、同所第8部主任研究員。



湯浦 克彦 (正会員)

昭和30年生。昭和53年東京大学工学部精密機械工学科卒業。昭和55年同大学院修士課程修了。同年より(株)日立製作所中央研究所に勤務。仮名漢字変換、自然言語処理のほかLISPを中心とするプログラミング言語の研究に従事。ソフトウェア学会会員。



武市 宣之 (正会員)

昭和18年生。昭和41年京都大学工学部電気工学第2学科卒業。昭和43年同大学院修士課程修了。同年(株)日立製作所入社。以来、同社中央研究所、システム開発研究所にて、教育システム、日本語ワードプロセッサ、仮名漢字変換、知識処理プログラミングの研究開発に従事。現在、同社中央研究所第8部主任研究員。電子通信学会会員。