2F-6

文庫本自動点字翻訳システムにおける漢字かな変換について

島田恭宏　塩野　充

岡山理科大学 工学部 電子工学科

1. まえがき

視覚障害者に対して文字情報の獲得を支援する為の，文庫本自動点字翻訳システムにおける漢字かな変換部分について述べる．現在，点字には，かな点字と漢点字があるが，漢点字においては，その構造も意味も違う2方式が存在し，かつ，かな点字と比較するとそれほど普及していないことから，本システムではかな点字を翻訳対象にした．このため一般に用いられる日本語文書（具体的な対象としては文庫本小説）が漢字かな混じり文であることから，漢字部分を読みに置き換える処理が必要となる．また，点字を表記する場合には分かち書きが必要となることから，これらの2つの課題を解決するための手法について考察する．

2. 変換用辞書辞書

漢字かな変換を行う場合，変換は単語単位で行う．ここでは変換用単語辞書について述べる．

2．1 漢字自立語辞書

漢字で始まる自立語を登録した辞書であり，その形式は見出し，読み，文法情報である．体言など活用の無い単語に関しては単語全てを見出し語とし，用言は，語幹のみを見出し語とした．ここで，上一段，および下一段活用については単に一段活用とすることとし，語幹に続く1文字も登録している．また漢字で始まる固有名詞，接辞なども一括して漢字自立語辞書に登録している．

2．2 ひらがな自立語辞書

ひらがなで始まる自立語を登録した辞書で，形式および登録語については漢字自立語辞書と同様である．この辞書はおもにひらがな表記される単語を中心に登録している．漢字表記する単語でも場合によっては（幼児用図書など）ひらがな表記をする場合があり，ある程度，漢字自立語辞書に登録した単語もひらがな自立語辞書に登録すべきであるが，今回作成した辞書にはこれらの単語は登録していない．ひらがな自立語辞書には漢字との混ぜ書き語も登録している．頻繁に見られる例としては女性の名前で「子」のみが漢字の場合などである．これら二つの自立語辞書は，パソコン用の日本語入力フロントエンドプロセッサの辞書等を基に作成した．

2．3 付属語辞書

助詞，助動詞に関して登録した辞書である．辞書形式は，前述の2つの辞書とは異なり，見出し語，前接続条件を記したテーブルのアドレス，活用形情報である．見出し語は助動詞に関しては各活用形に展開し，各々の活用形を異なる単語として別々に登録した．今回作成した辞書は付属語単独と，これら付属語が連接したものも1つの単語として登録している．前接続条件は，後続する付属語ごとにその直前の自立語，付属語，その活用形が決まるため，これらを記したテーブルを作成しその付属語に対応するテーブルのアドレスを記した．品詞活用形情報は，付属語の品詞，活用形を示している．付属語辞書は文献[1][2][3]を参考に作成した．

3. 変換アルゴリズム

前章で述べた単語辞書を用いて次の手順に従い変換を行う．

3．1 文節位置

漢字かな混じり文を文字種の移り変わる位置（ひらがな→他字種）によって文節単位に切る．この文節は処理を行う上での単位であり，長い文を一度に処理するのではなく，この文節を単位として処理を行う．単語照合を行う場合，最長一致の原則を用いているためその時間的効果は大きいと考える．

3．2 処理対象文節

原則的には処理対象は1つの文節であるが，場合によっては先の文節切りで1つの単語が文節に分けられてしまう場合がある．すなわち「成り立つ」という単語は「成り」と「立つ」に分割されてしまい，正確な単語照合ができない．よって，単語照合を行う場合，連接する2

On Kanji-Kana translation for Braille translation system for novel books.
Yasuhiro SHIMADA, Mitsuru SHIONO
Okayama University of Science.

つの文節を処理対象とし，最初の文節内において照合した単語の最長の単語が次の文節にかかるならば，その２つの文節を１つの文節と見なし，そうでなければ独立した２つの文節として以後の処理を行う．

３．３　照合

変換処理を行う上でまず必要なことは，照合すべき単語辞書の選択である．これは，文節ないし対象単語の先頭文字の字種によって行う．すなわち，先頭文字が漢字であれば漢字自立語辞書を，またひらがなであればひらがな自立語辞書を選択する．付属語辞書は，文節の構造が単純には＜自立部＋付属部＞となるので，文節内において最初の単語を照合した後，残りの文字列がひらがなから始まっている場合にのみ付属語辞書の照合を行う．但しこの場合，文節が自立語のみでも構成できるため，ひらがな自立語辞書の照合も行っておく．

３．４　文節構造検定

前述の辞書照合により，仮に決定された文節内において候補単語が複数挙がるわけであるが，これらを選択する際に，文節構造を調べることにより単語を選択する．文節構造は文献[4]等を参考にし，文節構造検定（単語構造，文節末語の条件，単語間の連接関係等）を行い文節となり得る条件を満たしている単語を選択，それにより得られる辞書に登録している読みと漢字を置き換えることによって漢字かな変換を行う．

分かち書きについては，点字の分かち書きの手法を規則化しこれに基づいて行う．点字の分かち書きの要約は次の通りである[5][6]．

(1)自立語は前を切って書く
(2)付属語は自立語に付けて書く
(3)合成語は構成単語単位に切って書く
(4)接頭語，および接尾語は自立語に付けて書く

漢字かな変換処理の処理の流れを図１に示す．各々の処理は再帰形関数（Ｆｕｎｃ）で構成しており，辞書照合の失敗，あるいは文節構造検定の失敗によってバックトラックが発生しても比較的処理しやすい構成をとっている[7]．

４．むすび

本稿では，点字翻訳システムにおける漢字かな変換について報告を行った．現在，表記のゆれに対応するために，単語辞書の整備を行っており，また文節構造検定における単語間の連接条件の検討を進めている．現在の手法のみでは，漢字列に対する適切な読み，同一品詞での同字異語に対する適切な読みが与えられないため，これらに対する検討も進めている．

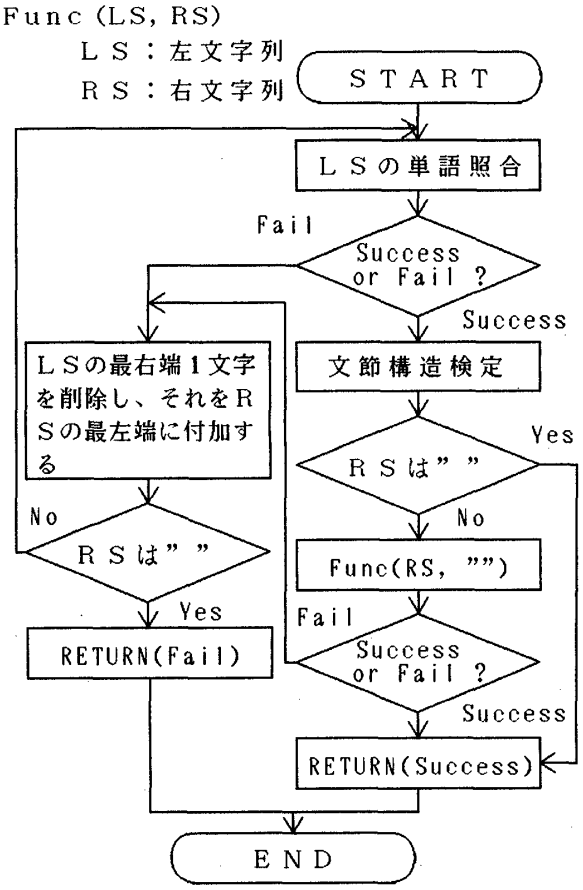

図１　処理の流れ

[参考文献]

（１）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４２，秀英出版，1979.

（２）遠藤嘉基監修：”対照日本文法（第72版）”，中央図書，1979.

（３）久松潜一，佐藤謙三編：”角川国語辞典（第164版）”，角川書店，1977.

（４）長尾　真監修：”日本語情報処理”，電子情報通信学会，コロナ社，1984.

（５）本間，岩橋，田中：”点字と朗読への招待”，福村出版，1983.

（６）野村典正，森　健一：”漢字かな変換システムの試作”，信学論，Vol. J66-D No. 7，1983.

（７）田中穂積，佐藤泰介，元吉文男：”自然言語処理のためのプログラミングシステム－拡張ＬＩＮＧＯＬについて－”，信学論，Vol. J60-D No12，1977.