

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2013-125399

(P2013-125399A)

(43) 公開日 平成25年6月24日(2013.6.24)

(51) Int.Cl.	F 1	テーマコード (参考)
G06F 17/22 (2006.01)	G06F 17/22 5 1 2 Z	5 B 0 9 1
G06F 17/27 (2006.01)	G06F 17/22 5 2 4 Z	5 B 1 0 9
	G06F 17/27 G	

審査請求 未請求 請求項の数 5 O L (全 15 頁)

(21) 出願番号 特願2011-273523 (P2011-273523)
 (22) 出願日 平成23年12月14日 (2011.12.14)

(71) 出願人 000002945
 オムロン株式会社
 京都市下京区塩小路通堀川東入南不動堂町
 801番地
 (74) 代理人 100078916
 弁理士 鈴木 由充
 (74) 代理人 100142114
 弁理士 小石川 由紀乃
 (72) 発明者 中山 拓也
 京都府京都市下京区塩小路通堀川東入南不
 動堂町801番地 オムロンソフトウェア
 株式会社内
 Fターム(参考) 5B091 AB13 EA01
 5B109 MA07 MG00

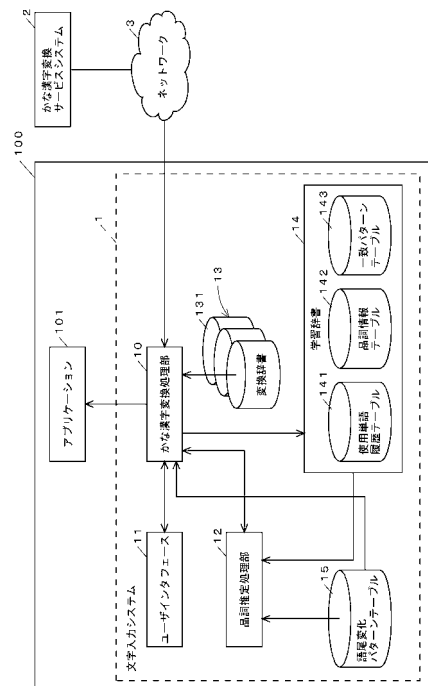
(54) 【発明の名称】 文字入力用のプログラムおよび情報処理装置

(57) 【要約】

【課題】品詞種別を特定できない語句を、正しい変換ができるように学習する。

【解決手段】ユーザの操作に応じてアプリケーション101への文字入力を行う文字入力システム1に、候補の選択によって確定した語句を品詞種別に対応づけて保存する学習辞書14と、複数の品詞種別がそれぞれの語尾変化のパターンと共に登録された語尾変化パターンテーブル15とを設ける。品詞種別を特定できない変換後文字列が確定された場合には、品詞推定処理部12が確定された文字列を語尾変化パターンテーブル15と照合し、当該文字列に整合する語尾変化のパターンを有する品詞種別と当該文字列の語尾とを推定する。またかな漢字変換処理部10は、推定が行われた変換後文字列とその語尾を除く部分に前方一致する変換後文字列との組み合わせ毎に語尾の集合を設定して一致パターンテーブル143内に保存し、この語尾の集合に適合する語尾変化のパターンを有する品詞種別を当該集合に対応する各変換後文字列に対応づける。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

変換前文字列の入力操作を受け付ける入力手段と、変換前文字列に対応する変換後文字列の候補を検索する検索手段と、検索手段により抽出された候補の中の1つを選択する操作に応じて選択された候補の変換後文字列を確定して出力する確定処理手段と、確定された変換後文字列を当該文字列に適合する品詞種別に対応づけて学習辞書に保存する学習処理手段とを備えた文字入力装置として、コンピュータを機能させるプログラムであって、

語尾が変化する品詞を語尾変化のパターンに基づき分類することにより設定された複数の品詞種別が前記語尾変化のパターンと共に登録された語尾変化パターンテーブルを前記文字入力装置に組み込むためのプログラムが含まれており、

前記学習処理手段は、品詞種別が特定されない変換後文字列を語尾変化パターンテーブルと照合して、当該変換後文字列に適合する語尾変化のパターンを有する品詞種別と当該変換後文字列の語尾とを推定する推定手段と、推定された語尾を保存するための記憶手段とを具備し、

前記推定手段は、品詞種別および語尾が推定された変換後文字列と当該変換後文字列の語尾を除く部分に前方一致する変換後文字列との組み合わせ毎に語尾の集合を作成して前記記憶手段に保存すると共に、推定された品詞種別のうち、前記語尾の集合に適合する語尾変化のパターンを有する品詞種別を当該集合に対応する各変換後文字列に対応づける、ことを特徴とする、文字入力用のプログラム。

【請求項 2】

前記語尾変化パターンテーブルには、さらに、語尾が変化しない品詞をその品詞の単語に添付され得る付属語のパターンに基づき分類することにより設定された品詞種別が、付属語のパターンを語尾変化のパターンとして少なくとも1つ登録されている、請求項 1 に記載された文字入力用のプログラム。

【請求項 3】

前記検索手段は、変換前文字列により学習辞書を検索すると共に、この検索により語尾変化パターンテーブルに登録されている品詞種別に対応づけられた変換後文字列を抽出したとき、その品詞種別に対応する語尾変化のパターンと抽出された変換後文字列の語頭部分とを組み合わせる複数種の候補を作成する、請求項 1 または 2 に記載された文字入力用のプログラム。

【請求項 4】

前記検索手段は、外部の変換システムに変換前文字列による変換処理を要求してその要求に対する当該変換システムからの応答により変換後文字列の候補を取得する機能を具備し、

前記推定手段は、外部の変換システムから取得した変換後文字列の候補が確定されたとき、その変換後文字列を前記語尾変化パターンテーブルと照合して品詞種別および語尾を推定し、当該変換後文字列を推定された語尾と当該語尾を除いた語頭部分とに区分けして、推定された品詞種別に対応づけて前記学習辞書に保存する、請求項 1 ~ 3 のいずれかに記載された文字入力用のプログラム。

【請求項 5】

変換前文字列の入力操作を受け付ける入力手段と、変換前文字列に対応する変換後文字列の候補を検索する検索手段と、検索手段により抽出された候補の中の1つを選択する操作に応じて選択された候補の変換後文字列を確定して装置内で起動中のアプリケーションに出力する確定処理手段と、確定された変換後文字列を当該文字列に適合する品詞種別に対応づけて学習辞書に保存する学習処理手段とを備えた情報処理装置であって、

語尾が変化する品詞を語尾変化のパターンに基づき分類することにより設定された複数の品詞種別が前記語尾変化のパターンと共に登録された語尾変化パターンテーブルがさらに設けられ、

前記学習処理手段は、品詞種別が特定されない変換後文字列を語尾変化パターンテーブルと照合して、当該変換後文字列に適合する語尾変化のパターンを有する品詞種別と当該

10

20

30

40

50

変換後文字列の語尾とを推定する推定手段と、推定された語尾を蓄積するための記憶手段とを具備し、

前記推定手段は、品詞種別および語尾が推定された変換後文字列と当該変換後文字列の語尾を除く部分に前方一致する変換後文字列との組み合わせ毎に語尾の集合を作成して前記記憶手段に蓄積すると共に、推定された品詞種別のうち、語尾の集合に適合する語尾変化のパターンを有する品詞種別を当該集合に対応する各変換後文字列に対応づける、ことを特徴とする、情報処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

10

本発明は、起動中のアプリケーションに文字を入力するための文字入力装置としてコンピュータを機能させるためのプログラム、文字入力機能を備えた情報処理装置に関する。

【背景技術】

【0002】

一般的な文字入力システム（IME）では、入力された変換前文字列によりシステム内の辞書を検索することによって変換後文字列の候補を複数抽出し、各候補のリストを表示して選択操作を受け付ける。また、利用される頻度が高い語句の呼び出しを容易にするために、選択により確定された変換後文字列を入力順序や品詞種別に対応づけて蓄積する学習辞書が設けられたシステムがある。さらに、変換の精度を向上するために、体言と付属語との組み合わせによる文字列を、語頭部分（体言）と語尾の部分（付属語）とを区分けして保存したり、動詞を表す文字列をその活用形と組み合わせで保存することが提案されている（特許文献1を参照）。

20

【0003】

さらに近年は、インターネットなどの通信システムを利用して、外部の変換システムに変換前文字列を送信して変換処理を行わせ、変換システムから送信された変換後文字列による候補を表示する機能を有する情報処理装置がある（たとえば、特許文献2の段落0075～0084，特許文献3の段落0047～0052を参照。）。以下、この機能を「外部変換機能」と呼び、当該機能により取得される候補を「外部変換候補」と呼ぶ。

【先行技術文献】

【特許文献】

30

【0004】

【特許文献1】特開平8-69459号公報

【特許文献2】特開2008-250606号公報

【特許文献3】特開2009-20865号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

外部変換機能によれば、大容量の辞書を持たない装置や辞書を全く持たない機器でも、変換前文字列に対して相当数の変換候補を取得することができる。また、新語を入力する場合にも容易に対応することができる。

40

また、学習辞書の仕組みを利用すれば、外部変換機能により取得した外部変換候補であっても、ユーザにより選択されて学習辞書に登録されたものであれば、次の入力から、携帯端末装置の変換機能によって呼び出すことができる。

【0006】

しかし、外部の変換システムにおいて採用している品詞の分類が装置で採用している分類と異なると、外部変換候補の品詞種別を正確に特定するのが困難になる。このため、選択された外部変換候補が誤った品詞種別に対応づけられて学習されると、その学習結果を利用した変換処理で誤った変換が行われるおそれがある。

【0007】

たとえば、動詞が名詞として認識された場合、その単語には付属語が付くと認識される

50

ので、単語自体の語尾を変化させることができず、後続の文字列との関係にも誤認識が生じる場合がある。たとえば、「ねむる」という変換前文字列に対して外部から取得した「眠る」という単語が名詞として学習されると、「ねむ・・・」という変換前文字列を「眠れない」や「眠らない」に変換することは不可能になる。また、「ねむるもり」という読み文字列に対する変換処理では、「も」が名詞の「ねむる」に対する係助詞であると判別されて、変換前文字列が「ねむる - もノリ」と切り分けられ、「眠るも利」のように変換される場合がある。

【0008】

本発明は上記の問題に着目し、外部のシステムから取得した単語など、品詞種別を一意に特定できない語句であっても、品詞種別を高い確度で推定して、正しい変換ができるように学習することを課題とする。

10

【課題を解決するための手段】

【0009】

本発明によるプログラムは、変換前文字列の入力操作を受け付ける入力手段と、変換前文字列に対応する変換後文字列の候補を検索する検索手段と、検索手段により抽出された候補の中の1つを選択する操作に応じて選択された候補の変換後文字列を確定して出力する確定処理手段と、確定された変換後文字列を当該文字列に適合する品詞種別に対応づけて学習辞書に保存する学習処理手段とを備えた文字入力装置として、コンピュータを機能させる。

【0010】

20

この文字入力装置には、語尾が変化する品詞を語尾変化のパターンに基づき分類することにより設定された複数の品詞種別が語尾変化のパターンと共に登録された語尾変化パターンテーブルが設けられる。

【0011】

学習処理手段は、品詞種別が特定されない変換後文字列を語尾変化パターンテーブルと照合して、当該変換後文字列に適合する語尾変化のパターンを有する品詞種別と当該変換後文字列の語尾とを推定する推定手段と、推定された語尾を保存するための記憶手段とを具備する。推定手段は、品詞種別および語尾が推定された変換後文字列と当該変換後文字列の語尾を除く部分（語頭部分）に前方一致する変換後文字列との組み合わせ毎に語尾の集合を作成して記憶手段に保存すると共に、推定された品詞種別のうち、語尾の集合に適合する語尾変化のパターンを有する品詞種別を当該集合に対応する各変換後文字列に対応づける。

30

【0012】

上記の構成によれば、動詞や形容詞などの語尾に変化が生じる単語が品詞種別を一意に特定できない状態で確定された場合、その単語を語尾変化パターンテーブルと照合することにより、品詞種別や語尾を推定することができる。ここで推定される品詞種別が1つに特定されない場合でも、同じ単語が語尾が異なる形態で確定されて複数種の語尾が保存されると、これらの語尾の集合に語尾変化のパターンが適合する品詞種別を対応づけの対象として絞り込むことができる。よって、絞り込まれた品詞種別の語尾パターンに基づき、読み文字列の切り分けや語尾を変形する処理を正確に行うことができるようになる。

40

【0013】

上記の語尾変化パターンテーブルには、さらに語尾が変化しない品詞をその品詞の単語に添付され得る付属語のパターンに基づき分類することにより設定された品詞種別を、付属語のパターンを語尾変化のパターンとして少なくとも1つ登録することもできる。この構成によれば、名詞のような語尾が変化しない品詞と語尾が変化する品詞との識別が容易になる。また語尾が変化しない品詞に属する単語を、その品詞に適合する付属語を付した形態に変換することが可能になる。

【0014】

また上記文字入力装置の一実施形態では、検索手段は、変換前文字列により学習辞書を検索すると共に、この検索により語尾変化パターンテーブルに登録されている品詞種別に

50

対応づけられた変換後文字列を抽出したとき、その品詞種別に対応する語尾変化のパターンと抽出された変換後文字列の語頭部分とを組み合わせることで複数種の候補を作成する。このようにすれば、学習辞書に保存されている語句に完全一致するものだけでなく、保存されている語句の末尾を変形させた文字列による候補を作成することが可能になる。

【0015】

上記文字入力装置の他の実施形態では、検索手段は、外部の変換システムに変換前文字列による変換処理を要求してその要求に対する当該変換システムからの応答により変換後文字列の候補を取得する機能を有する。また、推定手段は、外部の変換システムから取得した変換後文字列の候補が確定されたとき、その変換後文字列を語尾変化パターンテーブルと照合して品詞種別および語尾を推定し、当該変換後文字列を推定された語尾と当該語尾を除いた語頭部分とに区別して、推定された品詞種別に対応づけて学習辞書に保存する。

10

【0016】

上記の実施形態によれば、外部システムから取得して確定された変換後文字列の品詞情報が不明であっても、その文字列の品詞種別や語尾を推定し、その推定結果に基づき、変換後文字列の語頭部分と語尾とに区別して、品詞種別の推定結果と共に保存することができる。よって、保存された文字列と語頭部分が同一で語尾が異なる文字列が確定されると、保存された文字列の中から確定文字列に組み合わせるべき文字列を容易に見つけて、両者を組み合わせることができる。また組み合わせられた文字列に対応づけられている品詞種別を対象にして、品詞種別の絞り込みを効率良く行うことが可能になる。

20

【発明の効果】

【0017】

本発明によれば、文字入力装置で採用している品詞種別に適合しない変換後文字列がアプリケーションへの入力文字列として確定された場合でも、この変換後文字列に適合する可能性の高い品詞種別に対応づけて学習することができるので、学習された変換後文字列を利用した変換処理で誤った形態の文字列に変換させるのを防ぐことができる。よって、文字入力処理の精度や利便性を高めることができる。

【図面の簡単な説明】

【0018】

【図1】本発明が適用された文字入力システムの機能ブロック図である。

30

【図2】語尾変化パターンテーブルのデータ構成例を示す図である。

【図3】文字入力システムにおける処理の概略手順を示すフローチャートである。

【図4】学習辞書内の各テーブルに保存されるデータの構成およびテーブル間の関係を説明する図である。

【図5】学習辞書内の各テーブルに保存されるデータの構成およびテーブル間の関係を説明する図である。

【図6】予測候補リスト作成処理の詳細な手順を示すフローチャートである。

【図7】候補確定処理の詳細な手順を示すフローチャートである。

【発明を実施するための形態】

【0019】

40

図1は、本発明が適用された文字入力システムの構成例を示す。

この実施例の文字入力システム1は、携帯電話などの情報処理装置100内で、ユーザの操作に従って当該装置で起動中のアプリケーション101（メモ帳、メーラーなど）に文字を入力するためのものである。具体的に当該システム1には、かな漢字変換処理部10、ユーザインタフェース11、品詞推定処理部12、変換辞書13、学習辞書14、語尾変化パターンテーブル15などが含まれる。これらはいずれも、専用のプログラムを情報処理装置100に組み込むことによって設定される。

【0020】

変換辞書13には、開発者により選択された複数の単語の辞書データが登録される。辞書データには、単語の読み（変換前のかな文字列）および表記（変換後文字列）のほか、

50

品詞種別、優先度などの情報が含まれる。なお、この実施例の変換辞書 13 は、複数のデータファイル 131 により構成されるが、これに限らず、1つのデータファイルにまとめることも可能である。

【0021】

この実施例の品詞種別は、文法上の品詞より細かい単位に分類される。たとえば、動詞などの語尾が変化する品詞については、「カ行五段活用動詞」「ラ行五段活用動詞」「一段動詞」・・・というように、変化のパターン（活用形）に基づく分類により複数の品詞種別が設定される。また、名詞についても、添付され得る付属語のパターンの違いに基づく分類により、「普通名詞」「サ変名詞」などの品詞種別が設定される。ただし、付属語単体や感動詞など、分類されることなく、品詞毎に1つずつ品詞種別が設定されるものもある。

10

【0022】

学習辞書 14 には、ユーザの選択操作により確定されてアプリケーション 101 に入力された語句（以下、「確定語句」という。）が、入力された順に一定数まで蓄積される。具体的に学習辞書 14 には、確定語句を蓄積するための使用単語履歴テーブル 141 のほか、各確定語句に適合する品詞種別が保存される品詞情報テーブル 142、および品詞種別の特定に用いられる一致パターンが保存される一致パターンテーブル 143 が含まれる。これらのテーブル 141、142、143 の構成や格納されるデータについては、後で詳細に説明する。

【0023】

ユーザインタフェース 11 は、図示しない表示部に文字キーを含む操作部を立ち上げ、キー操作の内容をかな漢字変換処理部 10 に連絡しながら、かな漢字変換処理部 10 による処理（読み文字列の組み立て処理、候補リストの作成処理、候補の確定表示など）の結果を表示する。なお、操作部は画像によるものに限らず、実物のキーによる操作部を設けてもよい。

20

【0024】

かな漢字変換処理部 10 は、ユーザインタフェース 11 と協働して、読み文字列の組み立てや変換処理を行う。また、この実施例のかな漢字変換処理部 10 は、インターネットなどのネットワーク 3 を介して外部のかな漢字変換サービスシステム 2 に読み文字列を渡して変換処理を要求し、この要求に対するかな漢字変換サービスシステム 2 からの応答によって外部変換候補を取得することもできる。

30

【0025】

各種検索により抽出された候補はユーザインタフェース 11 によってリスト表示される。ユーザインタフェース 11 がいずれかの候補を選択する操作を受け付けると、かな漢字変換処理部 10 は、選択された候補の文字列を確定してアプリケーション 101 に出力すると共に、確定語句を学習辞書 14 に保存する。また、確定語句により学習辞書 14 を検索し、次に入力される可能性のある文字列を繋がり予測候補として抽出する。

【0026】

品詞推定処理部 12 は、かな漢字変換サービス 2 から提供された外部変換候補などの品詞種別を特定できない文字列を対象に、その品詞種別を推定する処理を実行する。語尾変化パターンテーブル 15 は、この推定処理のために、利用される頻度が高い品詞種別の語尾変化のパターンが種別毎に登録される。また、この実施例では、かな漢字変換処理部 10 が学習辞書 14 から読み出した候補の文字列から語尾が異なる他の文字列を候補として作成する処理（後記する予測候補リストの作成処理）でも、語尾変化パターンテーブル 15 が使用される。

40

【0027】

図 2 は、語尾変化パターンテーブル 15 内の登録情報の一例を示す。このテーブル 15 には、品詞種別毎の語尾変化パターンとして、その品詞に属する単語を含む語句の語尾のパターンの集合が格納される。各語尾パターンは、語頭を表す「～」記号と語尾の文字列との組み合わせにより表される。動詞のように、単語自体の末尾が変化する品詞種別では

50

、活用語尾を表す語尾パターンが設定される。一方、名詞のように、単語の末尾が変化しない品詞種別では、単語に添付され得る付属語を表す語尾パターンが設定される。

【0028】

図3は、上記の文字入力システム1で実行される処理の概略手順を示す。

この処理では、毎時のキー操作を受け付けながら(ステップS1)、その操作内容に応じた処理を実行する。

【0029】

各種操作に伴う処理を説明する。まず、読みを示す仮名が入力された場合(ステップS2が「YES」)には、変換前の読み文字列を組み立て(ステップS6)、当該読み文字列により学習辞書14を検索して予測候補リストを作成し(ステップS7)、このリスト内の各候補を表示する(ステップS11)。

10

【0030】

読み文字列が組み立てられている状態下で変換操作が行われた場合(ステップS3が「YES」)には、その読み文字列により変換辞書13および学習辞書14を検索して変換候補リストを作成し(ステップS8)、リスト内の各候補を表示する(ステップS11)。なお、ステップS8では、読み文字列に前方一致する語句を抽出し、抽出された語句の品詞種別に基づき読み文字列を語頭部分と語尾部分とに分けするなどして、変換後文字列の候補を作成する。さらに、必要に応じて、かな漢字変換サービスシステム2と通信を行って外部変換候補を取得する。なお、外部変換候補の取得は、再度の変換操作によって行われる場合もある。

20

【0031】

候補が表示された状態下でいずれかの候補が選択されると(ステップS4が「YES」)、選択された候補の文字列を確定する(ステップS9)。さらに確定された文字列(確定語句)に続く語句の候補リスト(繋がり予測リスト)を作成し(ステップS10)、リスト内の各候補を表示する(ステップS11)。

所定の時点で終了操作が行われると(ステップS5)、操作部の表示を消失させることにより文字入力処理を終了する。

【0032】

図3に示す処理の流れ自体は、従来のシステムによるものと同様である。しかし、太枠で示した予測候補リストの作成処理(ステップS7)と候補確定処理(ステップS9)には特有の特徴がある。以下、これらの処理に関連する構成や処理の手順を説明する。

30

【0033】

図4は、学習辞書14の各テーブル141, 142, 143の構成とこれらのテーブルに格納されるデータの内容を具体的に示す。

使用単語履歴テーブル141には、毎回の確定語句の読みおよび表記が確定された順序に従って格納される。読みおよび表記は、いずれも語頭部分と語尾部分とに分かれている(間にハイフン記号(-)が設けられる。)が、確定語句の内容によってはこの分けなしに保存される場合もある。その場合には確定語句全体が語頭部分とされ、語尾は「なし」とされる。

【0034】

各確定文字列には、さらに、使用頻度、使用順序、および2種類のコードが対応づけられている。使用頻度は、確定語句がこれまでに選択された頻度を示し、使用順序は、確定語句が一連の文字入力において何番目に入力されたかを示す。

40

【0035】

2種類のコードの一方は品詞情報テーブル142内の対応する情報を示し、他方のコードは一致パターンテーブル143内の対応する情報を示す。品詞情報テーブル142では、確定語句に適合する品詞種別がコード情報に組み合わせられて格納される。一致パターンテーブル143には、語尾変化パターンテーブル141に登録されている語尾パターンの中で確定語句の語尾に一致したパターン(以下、「一致パターン」という。)がコード情報に組み合わせられて格納される。

50

【0036】

品詞情報テーブル142や一致パターンテーブル143の情報は、確定語句と一対一に対応づけられるのではなく、語頭部分が一致する確定語句毎に共通の情報が保存される。また、品詞情報テーブル142には、全ての確定語句に対する情報が保存されるが、一致パターンテーブル143の情報は、最初の学習時に品詞情報を一意に特定できなかった確定語句、具体的には外部変換候補として取り込まれて確定された語句に対する一致パターンが保存される。

【0037】

以下、図2を合わせて参照しながら、図4の各テーブルの構成やテーブル間の関係を説明する。

図4中のA、Bは、アプリケーション101に入力された文のデータを、各確定語句をスラッシュ記号「/」により分けた形式にして示したものである。なお、この実施例では、1番目の『私は』は変換辞書13から抽出され、2番目の『走る』および3番目の『貴方も』は、かな漢字変換サービスシステム2から提供され、最後の『走った』は、学習辞書14に保存された『走る』のデータから作成されたものとする。

【0038】

文字入力処理が開始されて、図中のAに示す文『私は/走る』が入力された時点の各テーブルは、図中の(A1)(A2)(A3)のようになる。

1番目の確定語句の『私は』では、変換辞書13に普通名詞として登録されている「私」と整合したことにより導出されるので、検索の結果に基づき『私』と『は』とに区切られて、使用単語履歴テーブル141に保存される。また『私』に対応づけられている品詞種別の「普通名詞」が品詞情報テーブル142に書き込まれ、その書込先の「001」が使用単語履歴テーブル141に格納される。品詞情報が一意に特定されたので、一致パターンテーブル143に1番目の確定語句の一致パターンが保存されることはない。

【0039】

2番目の『走る』は、外部変換候補であったため、品詞種別を特定できない。このような文字列に対し、この実施例では、語尾変化パターンテーブル15を参照して、『走る』に整合する語尾パターンを含む品詞種別を抽出する。具体的には各語尾パターンによる後方一致検索を実施する。この検索は、図1の品詞推定処理部12により実施される。

【0040】

ここで図2を参照すると、語尾変化パターンテーブル15には、「～る」という語尾パターンが設定されている品詞種別が2つ(ラ行五段動詞と一段動詞)保存されている。よって上記の検索により、これら2つの品詞種別が抽出されて、共通のコード「002」に対応づけられて品詞情報テーブル142に保存される。また、一致した語尾パターンの「～る」が、コード「102」に対応づけられて一致パターンテーブル143に格納される。

【0041】

また上記の処理結果に基づき、確定語句『走る』は『走』と『る』とに区切られて使用単語履歴テーブル141に保存される。さらに、品詞情報テーブル142および一致パターンテーブル143内の対応情報のコード「002」「102」がこの確定語句に対応づけられて保存される。

【0042】

さらに、図4のBに示すように、『私は/走る』に続いて『貴方も/走った』という文が入力されると、各テーブル141、142、143は、図中の(B1)(B2)(B3)のようになる。

『貴方も』は、先の『走る』と同様に外部変換候補であるため、語尾変化パターンテーブル15から『貴方も』に整合する語尾パターンを含む品詞種別が抽出され、これが品詞情報テーブル142に保存される。再び図2を参照すると、語尾変化パターンテーブル15には、『貴方も』に整合する「～も」という語尾パターンを含む品詞種別が2つ(普通名詞およびサ変名詞)保存されているので、これら2つの品詞種別が抽出されて、共通の

10

20

30

40

50

コード「003」に対応づけられて品詞情報テーブル142に保存される。また、一致した語尾パターン「～も」が、コード「103」に対応づけられて一致パターンテーブル143に格納される。

【0043】

上記の処理結果に基づき、確定語句『貴方も』は『貴方』と『も』とに区切られて使用単語履歴テーブル141に保存される。さらに、品詞情報テーブル142および一致パターンテーブル143内の対応情報のコード「003」「103」がこの確定語句に対応づけられて保存される。

【0044】

4番目の確定文字列『走った』は、学習辞書14の検索により『走-る』の変化形として抽出されるので、この『走-る』に一致する語頭部分『走』と語尾の『った』とに分けられて使用単語履歴テーブル141に格納される。

10

【0045】

『走-った』に対しては、品詞情報テーブル142や一致パターンテーブル143への新たなコードは作成されず、『走-る』と同じコード(品詞情報テーブルのコード「002」と一致パターンテーブルのコード「102」)が保存される。ただし、『走-る』とは異なる語尾が付与されているので、この語尾パターン「～った」が一致パターンとして追加される。

【0046】

また、図2によれば、コード「102」に対応づけられた『～った』および『～る』が語尾変化パターンに含まれている品詞種別は「ラ行五段動詞」のみである。よって、品詞情報テーブル142内のコード002の情報中の「一段動詞」は削除される。

20

【0047】

図3に示した予測候補リストの作成処理(ステップS7)では、読み文字列による学習辞書14の検索により抽出された語句の語尾をそれぞれの品詞種別に基づき変化させることによって、複数の候補を作成する。また、変換候補リストの作成処理(ステップS8)では、変換辞書13や学習辞書14との照合により変換対象の読み文字列を語頭部分と語尾部分とに分けて、表記(かな漢字)をあてはめる。よって、辞書内の各単語の品詞種別が特定されていない場合には、誤変換が生じてしまう。

【0048】

しかし、この実施例では、品詞種別が特定できない文字列を学習辞書14に保存する際に、語尾変化パターンテーブル15との照合によって語頭部分と語尾部分とを識別し、当該文字列に該当する可能性のある品詞種別を推定することができる。よって、この推定された品詞種別に基づき、正しく変換された候補を抽出することが可能になる。

30

【0049】

たとえば、図4の例においては、4番目の『走った』の入力のために『はし』という読み文字列が入力されたとき、先に確定された『走る』に対して保存された品詞種別の「ラ行五段動詞」および「一段動詞」の各語尾パターンが『走』に適用される。よって、この段階では、『走ない』『走た』など、一段動詞の語尾変化パターンによる誤った変換候補も作成されるが、ラ行五段動詞の語尾変化パターンによる正しい変換候補も作成される。また、作成された候補の中から入力対象の『走った』が選択されると、『走』に対応する品詞種別はラ行五段動詞のみに絞り込まれるので、その後は、『走』を語頭とする語を入力する際に誤変換が生じることはなくなる。

40

【0050】

このように、かな漢字変換サービスシステム2から提供された外部変換候補の文字列についても、変換辞書13に登録されている単語と同じように、語尾を正しく変化させた変換候補を作成することが可能になる。また、動詞と名詞との識別など、広い概念での品詞種別の識別も可能になる。

【0051】

なお、品詞種別の推定の対象となる確定語句は外部変換候補に限らず、読み文字列から

50

確定されたひらがな文字列、変換処理により設定された語頭と語尾との区切りを変更する操作によって再変換されたかな漢字文字列、後変換操作に応じて変換されたカタカナ文字列やアルファベット文字列など、変換辞書 1 3 に登録されていない語句の全てを対象とすることができる。

【 0 0 5 2 】

また図 4 の例では、語尾変化パターンテーブル 1 5 との照合によって確定語句に整合する語尾パターンが 1 つに特定されたが、常にそうなるとは限らず、複数の語尾パターンが抽出される場合もある。その場合は、品詞情報テーブル 1 4 2 や一致パターンテーブル 1 4 3 には、抽出された語尾パターン毎に、異なるコードによる情報が保存される。

【 0 0 5 3 】

図 5 は、『ははは』という語の確定に伴って各テーブル 1 4 1 , 1 4 2 , 1 4 3 に保存される情報を示す。

この例の確定語句の『ははは』は、読み文字列がそのまま確定されたもので、品詞種別が不明であるため、語尾変化パターンテーブル 1 5 を用いて、確定語句に整合する語尾パターンを探す処理が実施される。図 2 によれば、確定語句『ははは』に整合する変化パターンは普通名詞およびサ変名詞の『～は』および『～(付属語なし)』の 2 つになる。

【 0 0 5 4 】

よって、使用単語履歴テーブル 1 4 1 には、第 1 の語尾パターン『～は』に基づく『はは - は』と分けられた形態の語と、第 2 の語尾パターン『～(付属語なし)』に基づき『ははは』全体を語頭とする形態の語とが保存される。また、前者の形態に関しては、品詞情報テーブル 1 4 2 に『～は』を語尾変化パターンに含む「普通名詞」および「サ変名詞」がコード 0 0 4 に対応づけられて保存され、一致パターンテーブル 1 4 3 には、『～は』がコード 1 0 4 に対応づけられて保存される。後者の形態に関しては、品詞情報テーブル 1 4 2 に、『～』を語尾変化パターンに含む「普通名詞」「サ変名詞」「感動詞」の 3 種類の品詞種別がコード 0 0 5 に対応づけられて保存され、一致パターンテーブルに、『～』がコード 1 0 5 に対応づけられて保存される。使用単語履歴テーブル 1 4 1 では、『はは - は』と『ははは』とが、それぞれテーブル 1 4 2 , 1 4 3 内の対応するコードに組み合わせられて保存される。

【 0 0 5 5 】

品詞種別の推定の対象となる語句は、上記のほかにも考えられる。

たとえば、古い文字入力システムを残して新しい文字入力システムを導入して使用する場合に、古いシステムの辞書から読み出した単語を入力対象として確定したり、新語などが格納された変換辞書をシステムに追加してその新しい変換辞書から読み出した単語を入力対象として確定する場合にも、上記実施例と同様の方法により、確定語句の品詞種別を推定したり、学習された語句の品詞種別を絞り込むことができる。

【 0 0 5 6 】

つぎに図 6 を参照して、予測候補リストの作成処理(図 3 のステップ S 7)の詳細な処理手順を説明する。

この処理では、最初のステップ S 1 0 1 で、予測候補リストをクリアする。以下に述べるように、この予測候補リストは 2 つに分かれており、ステップ S 1 0 2 で一方のリストが作成され、ステップ S 1 0 3 ~ S 1 0 7 により、他方のリストが作成される。

【 0 0 5 7 】

ステップ S 1 0 2 では、変換辞書 1 3 に登録されている単語を対象として、読み文字列による前方一致検索を実行し、抽出された候補により第 1 の予測候補リストを作成する。

【 0 0 5 8 】

第 2 の予測候補リストの作成では、まず学習辞書 1 4 に保存されている語句の語頭部分を対象にして、前方一致検索(語頭部分が読み文字列に包含されるものを抽出)を実行する。以下、抽出された語句を順に対象として、ステップ S 1 0 5 , S 1 0 6 , S 1 0 7 を実行する。

【 0 0 5 9 】

10

20

30

40

50

ステップ S 1 0 5 では、ステップ S 1 0 3 の検索で一致した語頭部分を抽出する。ステップ S 1 0 6 では、処理対象の語句に対応づけられている品詞種別の語尾パターンを全て読み出す。ここでは、複数の品詞種別に対応づけられている場合には、全ての品詞種別の語尾パターンが読み出される。

ステップ S 1 0 7 では、読み出された各語尾パターンを、それぞれステップ S 1 0 5 で抽出した語頭部分に組み合わせることによって、複数の候補の文字列を作成する。さらにこれらの候補を第 2 の予測候補リストに保存する。

【 0 0 6 0 】

ステップ S 1 0 3 で抽出された語句の全てに対する処理が終了すると（ステップ S 1 0 4 が「YES」）、ステップ S 1 0 8 に進み、第 1 および第 2 の予測候補リストを統合する。たとえば、双方のリスト間で重複している候補を 1 つにまとめたり、各候補に優先順位を設定する処理などが行われる。統合されたリストは、図 3 のステップ S 1 1 の処理により、表示される。

10

【 0 0 6 1 】

つぎに図 7 を参照して、候補の選択操作に伴う候補確定処理（図 3 のステップ S 9 ）の詳細な手順を説明する。

【 0 0 6 2 】

この処理では、まず、選択された候補の文字列（確定語句）により学習辞書 1 4 の使用単語履歴テーブル 1 4 1 を検索して、読みおよび表記が共に確定語句に前方一致する語句を抽出する。なお、確定語句が変換辞書 1 3 や学習辞書 1 4 の検索により抽出された場合には、その検索の際に語頭と語尾とに区別されているので、ステップ S 2 0 1 では、語頭部分が一致する語句を検索する。

20

【 0 0 6 3 】

上記の検索で確定語句に前方一致する語句が抽出されなかった場合（ステップ S 2 0 2 が「NO」）には、ステップ S 2 0 3 および S 2 0 4 を実行する。

ステップ S 2 0 3 では、語尾変化パターンテーブル 1 5 から確定語句に整合する語尾パターンを有する品詞種別を抽出する。また、抽出された品詞種別を品詞情報テーブル 1 4 2 に保存する。さらに、ステップ S 2 0 4 では、確定語句に整合した語尾パターンを一致パターンテーブル 1 4 3 に保存する。

【 0 0 6 4 】

30

上記の品詞種別や一致パターンは、いずれもコードに対応づけて保存される。これらの保存が終了すると、ステップ S 2 0 9 に進み、確定語句に関する諸情報（上記のコードを含む）を使用単語履歴テーブル 1 4 1 に保存する。

この後は、読み文字列をクリアする処理（ステップ S 2 1 0 ）、および確定語句をアプリケーションに出力する処理（ステップ S 2 1 1 ）を実行し、処理を終了する。

【 0 0 6 5 】

ステップ S 2 0 1 の検索で確定語句に前方一致する語句が抽出されたが、その語句に複数の品詞種別に対応づけられている場合（ステップ S 2 0 2 が「YES」、ステップ S 2 0 5 が「NO」）には、抽出された語句に対応づけられているコードにより、抽出された語句に対応する品詞種別および一致パターンを特定する（ステップ S 2 0 6 ）。これらは、つぎのステップ S 2 0 7 およびステップ S 2 0 8 の処理対象となる。

40

【 0 0 6 6 】

ステップ S 2 0 7 では、確定語句の語尾パターンを処理対象の一致パターンとして追加する。ステップ S 2 0 8 では、追加後の一致パターンのリストと処理対象の品詞種別の語尾変化パターンとを照合し、一致パターンのリストに整合する語尾変化パターンを有する品詞種別を残し、整合しない品詞種別を削除する（ステップ S 2 0 8 ）。この後は、前述のステップ S 2 0 9 に進み、確定語句に関する諸情報を使用単語履歴テーブル 1 4 1 に保存する。この保存情報には、ステップ S 2 0 6 で特定された品詞種別および一致パターンのコードが含まれる。

なお、確定語句の語尾のパターンが既に一致パターンに含まれている場合には、ステッ

50

ブ S 2 0 7 , S 2 0 8 はスキップされる。

【 0 0 6 7 】

ステップ S 2 0 1 の検索処理で確定語句に前方一致する語句が抽出され、その語句の品詞種別が確定されている場合 (ステップ S 2 0 2 , S 2 0 5 が「 Y E S 」) には、品詞種別や一致パターンを推定する必要がないので、すぐにステップ S 2 0 9 に進む。この場合のステップ S 2 0 9 では、確定語句の読みおよび表記を、当該確定語句を候補として導出した際に識別した語頭と語尾とに分けて保存すると共に、ステップ S 2 0 1 の検索によりヒットした語句と同じ品詞種別を対応づける。

さらに、読み文字列のクリア (ステップ S 2 1 0) や確定語句をアプリケーション 1 0 1 に出力する処理 (ステップ S 2 1 1) を実行し、処理を終了する。

10

【 0 0 6 8 】

なお、外部変換候補として取り込まれて学習辞書 1 4 に保存された単語であっても、変換辞書 1 3 のアップロードなどにより変換辞書 1 3 に登録される場合がある。この場合には、学習辞書 1 4 内の当該単語の品詞種別も変換辞書 1 3 に適合するように更新される。

【 0 0 6 9 】

上記の処理によれば、品詞種別が不明または品詞種別が一意に特定できない単語でも、語尾パターンが様々な形態に変更されて複数回の入力が行われるうちに、品詞種別を絞り込むことができる。よって、変換辞書 1 3 に登録されている単語と同様に、ステップ S 2 0 5 が「 Y E S 」となり、学習処理が容易になる。

【 0 0 7 0 】

20

ただし、図 7 のステップ S 2 0 8 において、一致パターンのリストの全てに一致する語尾変化パターンのみが残るように品詞種別を絞り込むと、ユーザが候補の選択を誤った場合に、以下のような支障が生じるおそれがある。

【 0 0 7 1 】

たとえば、図 4 の例では、『走る』がラ行五段動詞および一段動詞に対応づけられて保存された後に、『走った』を入力するために『はし』という読み文字列が組み立てられたとき、ラ行五段動詞の語尾変化パターンによる予測変換候補 (『走った』など) のほか、一段動詞の語尾変化パターンによる予測変換候補 (『走た』など) も表示される。これらの候補の中から、ユーザが誤って『走た』を選択すると、『走』に関する品詞種別は一段動詞に絞り込まれ、ラ行五段動詞との対応関係が解消されてしまう。そうなると、『走』

30

については、『走る』以外の活用形を正しく導出することが不可能になる。

【 0 0 7 2 】

この問題を重視する場合には、ステップ S 2 0 8 での品詞種別の削除の条件をやや緩和するのが望ましい。たとえば、一致パターンのリストの全てが一致しない語尾変化パターンを有する品詞種別でも、不一致の数があらかじめ定めた許容値までであれば、削除せずに残すようにするとよい。または一致パターンのリストに含まれるパターンの数が所定数に達するまでは品詞種別を絞り込みせずに、品詞種別毎にその種別に対応する一致パターンに整合する語句が選択された頻度を計数し、一致パターンの数が所定数に達してから、各品詞種別の選択頻度に基づいて、いずれの品詞種別を残すかを決めてもよい。

これらの処理によれば、候補の選択操作に多少の誤りがあった場合でも、正しい品詞種別が削除されてしまうのを防ぐことができる。

40

【 0 0 7 3 】

なお、上記の実施例では、学習辞書 1 4 内に一致パターンテーブル 1 4 3 を設けて、このテーブル 1 4 3 により、品詞種別を一意に特定できない語句を語頭部分が共通するもの毎に組にして、組毎に一致パターンを保存したが、語尾パターンの蓄積場所は学習辞書 1 4 には限らず、バッファに一時保存するに留め、品詞種別が 1 つに絞り込まれた語句に対しては一致パターンを消去してもよい。一致パターンを学習辞書 1 4 内に保存する場合も同様に、品詞種別が 1 つに絞り込まれた語句に対する情報を削除してもよい。

【 0 0 7 4 】

また、上記の実施例では、品詞情報テーブル 1 4 2 および一致パターンテーブル 1 4 3

50

の保存情報にコードを含めて、このコードを介して各テーブル142, 143の情報と使用単語履歴テーブル141に保存される語句とを対応づけたが、これに限らず、品詞種別や一致パターンの格納先のアドレスを介した対応づけを行ってもよい。また、品詞情報テーブル142を設けずに、使用単語履歴テーブル141の各レコードに品詞種別を含めるようにしてもよい。

【0075】

最後に、上記の文字入力システム1は、携帯型の情報処理装置に限らず、据え置き型の装置(ディスプレイ装置やファクシミリなど)に組み込むこともできる。またパーソナルコンピュータにも適用することが可能である。また、入力対象の言語は日本語に限らず、英語などの他の言語の入力にも、同様の構成のシステムにより対応することができる。

10

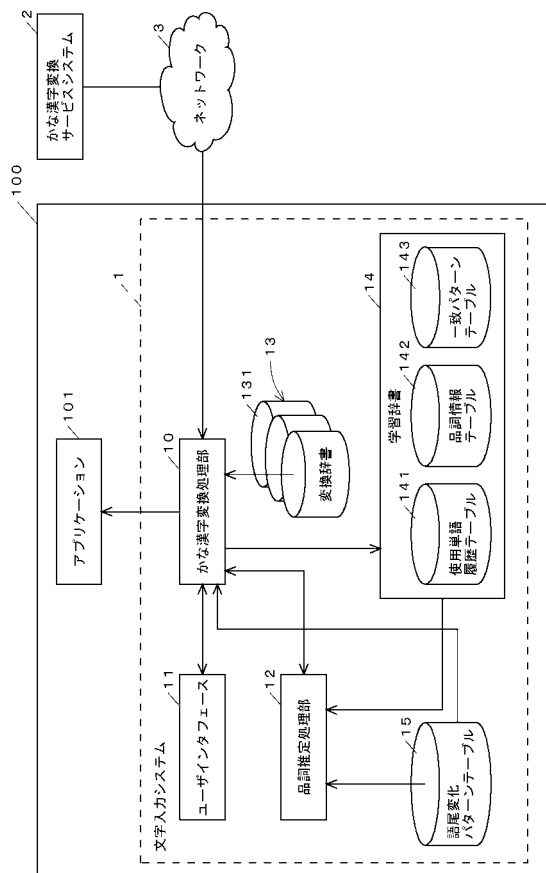
【符号の説明】

【0076】

- 1 文字入力システム
- 2 かな漢字変換サービスシステム
- 10 かな漢字変換処理部
- 11 ユーザインタフェース
- 12 品詞推定処理部
- 13 変換辞書
- 14 学習辞書
- 15 語尾変化パターンテーブル
- 101 アプリケーション
- 141 使用単語履歴テーブル
- 142 品詞情報テーブル
- 143 一致パターンテーブル

20

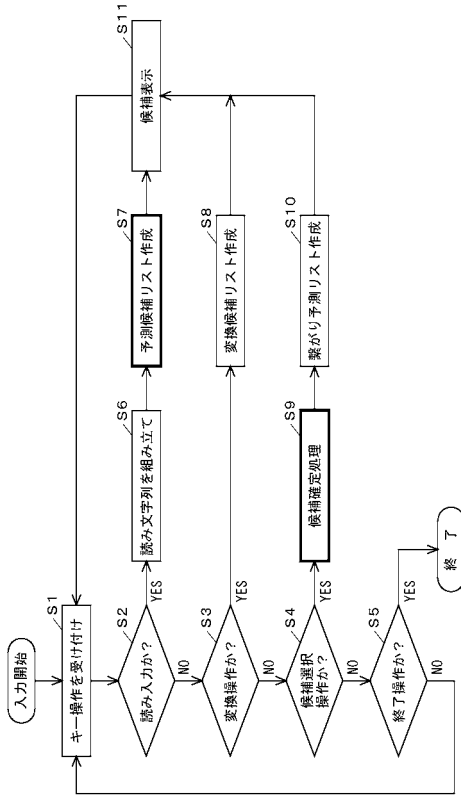
【図1】



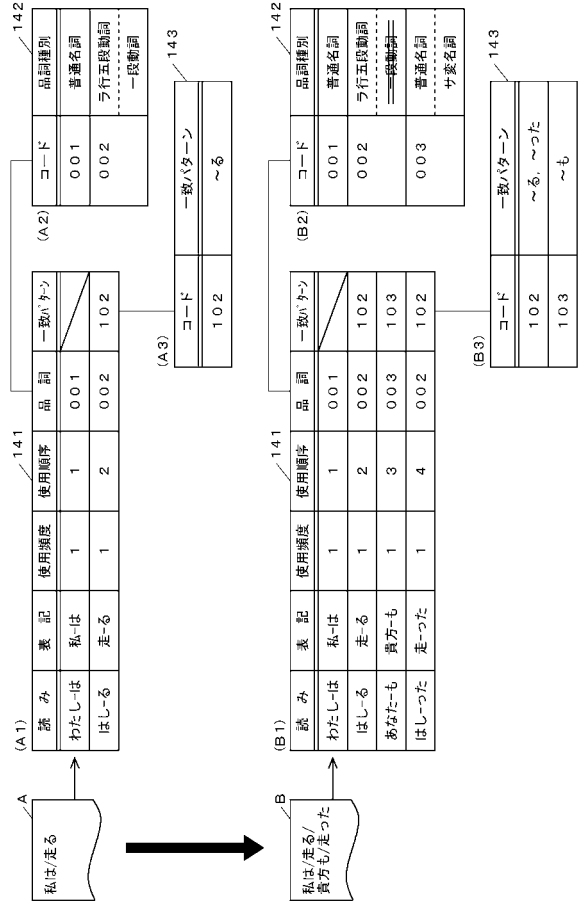
【図2】

品詞種別	語尾変化パターン
ラ行五段動詞	～らない, ～る, ～り, ～った, ～って, ～れ, ～ろう
バ行五段動詞	～ばない, ～ぶ, ～び, ～んだ, ～んで, ～べ, ～ぼう
一段動詞	～ない, ～る, ～ます, ～た, ～て, ～る, ～よう
普通名詞	～(付属語なし), ～は, ～が, ～を, ～で, ～の, ～に, ～も
サ変名詞	～(付属語なし), ～は, ～が, ～を, ～で, ～の, ～に, ～も, ～する
感動詞	～(付属語なし)
...	...

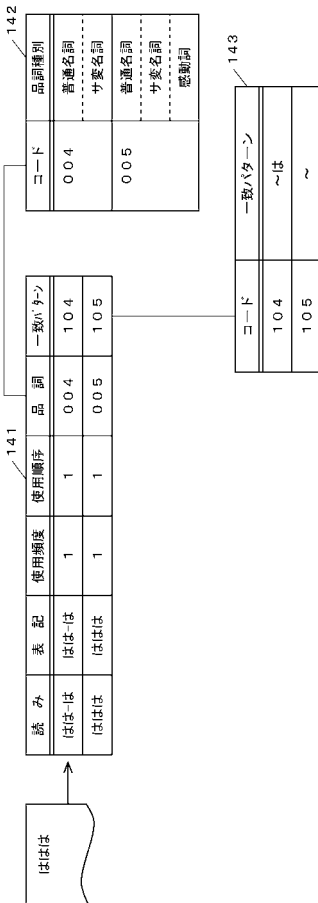
【図 3】



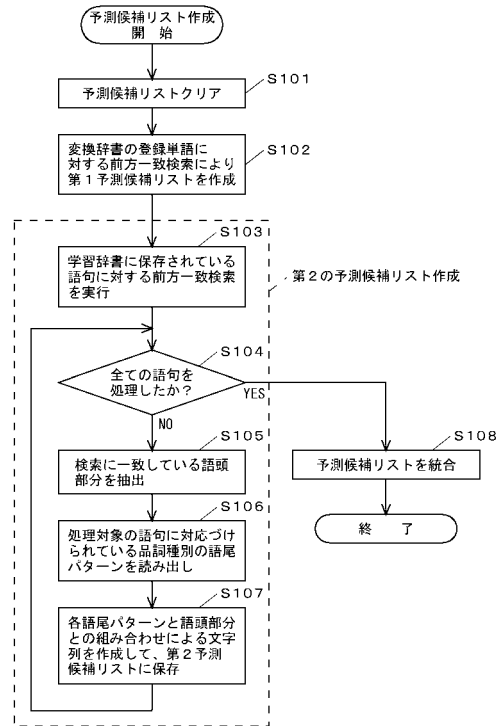
【図 4】



【図 5】



【図 6】



【 図 7 】

