

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6244993号  
(P6244993)

(45) 発行日 平成29年12月13日 (2017.12.13)

(24) 登録日 平成29年11月24日 (2017.11.24)

(51) Int.Cl.		F 1			
<b>G 0 6 F</b>	<b>17/28</b>	<b>(2006.01)</b>	<b>G 0 6 F</b>	<b>17/28</b>	<b>6 6 3</b>
<b>G 0 6 F</b>	<b>17/22</b>	<b>(2006.01)</b>	<b>G 0 6 F</b>	<b>17/22</b>	<b>6 7 6</b>

請求項の数 7 (全 18 頁)

(21) 出願番号	特願2014-45580 (P2014-45580)	(73) 特許権者	000005223
(22) 出願日	平成26年3月7日 (2014.3.7)		富士通株式会社
(65) 公開番号	特開2015-170218 (P2015-170218A)		神奈川県川崎市中原区上小田中4丁目1番1号
(43) 公開日	平成27年9月28日 (2015.9.28)	(74) 代理人	100099759
審査請求日	平成28年11月2日 (2016.11.2)		弁理士 青木 篤
		(74) 代理人	100119987
			弁理士 伊坪 公一
		(74) 代理人	100081330
			弁理士 樋口 外治
		(72) 発明者	高橋 潤
			神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		審査官	成瀬 博之

最終頁に続く

(54) 【発明の名称】 拗促音判定装置、拗促音判定方法及び拗促音判定用コンピュータプログラム

(57) 【特許請求の範囲】

【請求項 1】

少なくとも一つの漢字を含む語句の表記を表す表記データ及び当該語句の読みを拗音及び促音と直音とを区別せずに表す読みデータを取得する取得部と、

複数の漢字のそれぞれの読み及び当該読みが音読みか訓読みかを表す音訓種別が登録された単漢字辞書を記憶する記憶部と、

前記単漢字辞書を参照して、前記表記データに表された前記語句に含まれる前記少なくとも一つの漢字のそれぞれの読みの候補を検出する読み候補検出部と、

前記読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出する拗促音候補抽出部と、

前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定し、該特定された候補に対応する前記語句の読みの部分を、当該漢字の読みとする対応付け部と、

前記語句に含まれる前記少なくとも一つの漢字のうちの前記拗促音候補と対応する第1の漢字の読みにおける前記拗促音候補の位置、及び、前記第1の漢字の前記読みの音訓種別に応じて前記拗促音候補を拗音及び促音の何れか否か判定する判定部と、  
を有し、

前記判定部は、前記拗促音候補が促音となり得る音である場合、前記第1の漢字と前記第1の漢字に後続する漢字の組み合わせに対して、互いに独立して用いられる単語間の境

10

20

界を検出する所定の手法を適用することで、前記第 1 の漢字と前記第 1 の漢字に後続する漢字との間が前記境界であるか否か判定し、前記第 1 の漢字の前記読みにおける前記拗促音候補の位置、前記拗促音候補が前記境界と接しているか否か、及び、前記第 1 の漢字の前記読みの音訓種別に応じて、前記拗促音候補が促音か否かを判定する拗促音判定装置。

【請求項 2】

前記判定部は、前記拗促音候補が、促音となり得る音であり、前記第 1 の漢字の前記読みの末尾に位置し、かつ、前記境界と接しておらず、かつ、前記第 1 の漢字の前記読みが音読みである場合、前記拗促音候補を促音とする、請求項 1 に記載の拗促音判定装置。

【請求項 3】

前記判定部は、前記拗促音候補が、拗音となり得る音であり、かつ、前記第 1 の漢字の読みの先頭以外に位置し、かつ、前記第 1 の漢字の前記読みが音読みである場合、前記拗促音候補を拗音とする、請求項 1 または 2 に記載の拗促音判定装置。

10

【請求項 4】

前記対応付け部は、前記拗促音候補のうち、促音となり得る音を、促音に置換される可能性がある音に置換して、前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定する、請求項 1 ～ 3 の何れか一項に記載の拗促音判定装置。

【請求項 5】

前記対応付け部は、前記漢字の読みの候補に含まれる拗音及び促音が、当該拗音及び促音を直音化した音と一致するとみなして、前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定する、請求項 1 ～ 4 の何れか一項に記載の拗促音判定装置。

20

【請求項 6】

コンピュータが、複数の漢字のそれぞれの読み及び当該読みが音読みか訓読みかを表す音訓種別が登録された単漢字辞書を参照して、少なくとも一つの漢字を含む語句の表記を表す表記データに表された前記語句に含まれる前記少なくとも一つの漢字のそれぞれの読みの候補を検出し、

前記コンピュータが、拗音及び促音と直音とを区別せずに表す読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出し、

30

前記コンピュータが、前記読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出し、

前記コンピュータが、前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定し、該特定された候補に対応する前記語句の読みの部分を、当該漢字の読みとし、

前記コンピュータが、前記語句に含まれる前記少なくとも一つの漢字のうちの前記拗促音候補と対応する第 1 の漢字の読みにおける前記拗促音候補の位置、及び、前記第 1 の漢字の前記読みの音訓種別に応じて前記拗促音候補を拗音及び促音の何れか否か判定する

40

ことを含み、

前記拗促音候補を拗音及び促音の何れか否か判定することは、前記拗促音候補が促音となり得る音である場合、前記第 1 の漢字と前記第 1 の漢字に後続する漢字の組み合わせに対して、互いに独立して用いられる単語間の境界を検出する所定の手法を適用することで、前記第 1 の漢字と前記第 1 の漢字に後続する漢字との間が前記境界であるか否か判定し、前記第 1 の漢字の前記読みにおける前記拗促音候補の位置、前記拗促音候補が前記境界と接しているか否か、及び、前記第 1 の漢字の前記読みの音訓種別に応じて、前記拗促音候補が促音か否かを判定することを含む拗促音判定方法。

【請求項 7】

50

複数の漢字のそれぞれの読み及び当該読みが音読みか訓読みかを表す音訓種別が登録された単漢字辞書を参照して、少なくとも一つの漢字を含む語句の表記を表す表記データに表された前記語句に含まれる前記少なくとも一つの漢字のそれぞれの読みの候補を検出し、

拗音及び促音と直音とを区別せずに表す読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出し、

前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定し、該特定された候補に対応する前記語句の読みの部分を、当該漢字の読みとし、

前記語句に含まれる前記少なくとも一つの漢字のうちの前記拗促音候補と対応する第1の漢字の読みにおける前記拗促音候補の位置、及び、前記第1の漢字の前記読みの音訓種別に応じて前記拗促音候補を拗音及び促音の何れかが否か判定する、  
ことをコンピュータに実行させ、

前記拗促音候補を拗音及び促音の何れかが否か判定することは、前記拗促音候補が促音となり得る音である場合、前記第1の漢字と前記第1の漢字に後続する漢字の組み合わせに対して、互いに独立して用いられる単語間の境界を検出する所定の手法を適用することで、前記第1の漢字と前記第1の漢字に後続する漢字との間が前記境界であるか否か判定し、前記第1の漢字の前記読みにおける前記拗促音候補の位置、前記拗促音候補が前記境界と接しているか否か、及び、前記第1の漢字の前記読みの音訓種別に応じて、前記拗促音候補が促音か否かを判定する

ことを含む拗促音判定用コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば、語句の読みに含まれる音が拗音または促音であるか否かを判定する拗促音判定装置、拗促音判定方法及び拗促音判定用コンピュータプログラムに関する。

【背景技術】

【0002】

近年、音声で情報を提供するサービスが利用されている。このようなサービスを提供する装置は、例えば、提供対象となる情報を記した文章を表すテキストデータに対して音声合成処理を適用することで、その情報の音声データを生成する。

【0003】

提供対象となる情報の音声データを生成するために、その情報を記した文章中に登場する各語句を漢字などを用いて表した「表記」に対する、その語句の音声を表す「読み」を記述した辞書が用いられる。この辞書に登録されていない語句が、入力されたテキストデータまたは音声に含まれていると、その語句の読みが不明であるために、読みが誤った合成音声が生じることがある。そこで、辞書には、できるだけ多くの語句が登録されていることが好ましい。例えば、このような辞書には、数万個～数十万個の語句が登録される。

【0004】

そのため、新規に辞書を作成する際、作業者が手作業で語句を辞書に登録しようとする、その作業量は膨大となる。そこで、語句の表記とともにその語句の読みの情報を含む既存のデータベース、例えば、様々なWebページなどに対して、コンピュータにより様々な語句を抽出し、抽出した語句の表記と読みを辞書に自動的に登録することが好ましい。

【0005】

しかしながら、既存のデータベースには、「ャ」、「ュ」、「ョ」といった拗音、及び「ッ」といった促音を文字コードとして利用できず、拗音及び促音が直音と区別されず、例えば、拗音または促音を含む読みが全て大文字の仮名で表記されたものがある。このようなデータベースを辞書作成に利用するデータベースから排除するには、作業者が、例え

10

20

30

40

50

ば、目視で読み表記を確認する作業が必要となり、煩雑である。また、辞書作成に利用できるデータベースが限定されるので、結果的に辞書に登録されない語句の数が増えるおそれがある。そこで、拗音及び促音が正しく表記されていないデータベースも、辞書作成に利用できることが好ましい。

【 0 0 0 6 】

一方、入力された文字列の中に含まれる拗促音候補の文字を拗促音化可能か否かを判断し、可能と判断した場合に拗音化または促音化された文字列を表示する技術が提案されている（例えば、特許文献 1 を参照）。特許文献 1 に記載の技術は、拗音候補直前文字と拗音化可能文字の組み合わせを表すテーブルなどを参照することで、拗促音候補の文字を拗促音化可能か否か判定する。

10

【先行技術文献】

【特許文献】

【 0 0 0 7 】

【特許文献 1】特開平 6 - 4 5 2 0 号公報

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 8 】

しかし、特許文献 1 に記載の技術では、入力された文字列中において拗促音化可能な部分が表示されるにすぎず、結局、作業者が、その部分を拗促音化するか否かを最終的に判断する必要がある。

20

【 0 0 0 9 】

そこで本明細書は、一つの側面として、語句の読みのうちで拗音化または促音化する音を適切に判定できる拗促音判定装置を提供することを目的とする。

【課題を解決するための手段】

【 0 0 1 0 】

一つの実施形態によれば、拗促音判定装置が提供される。この拗促音判定装置は、少なくとも一つの漢字を含む語句の表記を表す表記データ及びその語句の読みを拗音及び促音と直音とを区別せずに表す読みデータを取得する取得部と、複数の漢字のそれぞれの読み及びその読みの音訓種別が登録された単漢字辞書を記憶する記憶部と、単漢字辞書を参照して、表記データに表された語句に含まれる少なくとも一つの漢字のそれぞれの読みの候補を検出する読み候補検出部と、読みデータに表された語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出する拗促音候補抽出部と、表記データに表された語句に含まれる少なくとも一つの漢字のそれぞれについて、その漢字の読みの候補のうち、その語句の読みと最も一致する候補を特定し、特定された候補に対応するその語句の読みの部分を、その漢字の読みとする対応付け部と、表記データに表された語句に含まれる少なくとも一つの漢字のうちの拗促音候補と対応する第 1 の漢字の読みにおける拗促音候補の位置、及び、第 1 の漢字の読みの音訓種別に応じて拗促音候補を拗音及び促音の何れかが否か判定する判定部とを有する。

30

【 0 0 1 1 】

本発明の目的及び利点は、請求項において特に指摘されたエレメント及び組み合わせにより実現され、かつ達成される。

40

上記の一般的な記述及び下記の詳細な記述の何れも、例示的かつ説明的なものであり、請求項のように、本発明を限定するものではないことを理解されたい。

【発明の効果】

【 0 0 1 2 】

本明細書に開示された拗促音判定装置は、語句の読みのうちで拗音化または促音化する音を適切に判定できる。

【図面の簡単な説明】

【 0 0 1 3 】

50

【図 1】拗促音判定装置の一例である、辞書登録装置の概略構成図である。

【図 2】処理部の機能ブロック図である。

【図 3】拗音を含む可能性がある音の並びのパターンを示す図である。

【図 4】促音を含む可能性がある音の並びのパターンを示す図である。

【図 5】拗音候補と対応漢字の読みとの位置関係の一例を示す図である。

【図 6】促音候補と対応漢字の読み及び単語境界との位置関係の一例を示す図である。

【図 7】共起確率データベースの一例を示す図である。

【図 8】拗促音判定処理を含む辞書登録処理の動作フローチャートである。

【図 9】変形例による処理部の機能ブロック図である。

【発明を実施するための形態】

10

【0014】

以下、図を参照しつつ、拗促音判定装置について説明する。

発明者は、拗音化または促音化される音は、特定の音の並びに含まれるとともに、その音に対応する漢字の読みにおける位置に関して特定の条件を満たすことに着目した。そこでこの拗促音判定装置は、拗音及び促音が直音と区別されることなく表された語句の読みを表す読みデータを取得する。そしてこの拗促音判定装置は、着目する語句の読みに含まれる、拗音または促音となり得る音のうち、拗音または促音を含む可能性がある特定の音の並びに含まれるとともに、対応する漢字の読みにおける位置に関する条件を満たすものを、拗音化または促音化する。

【0015】

20

図 1 は、拗促音判定装置の一例である、辞書登録装置の概略構成図である。本実施形態では、辞書登録装置 1 は、通信部 2 と、記憶部 3 と、表示部 4 と、処理部 5 とを有する。通信部 2、記憶部 3 及び表示部 4 は、処理部 5 とバスを介して接続されている。

【0016】

通信部 2 は、通信ネットワークに辞書登録装置 1 を接続するためのインターフェース回路を有する。通信部 2 は、取得部の一例であり、登録対象となる語句の表記データ及び読みデータを通信ネットワークを介して辞書登録装置 1 と接続された他の機器から取得する。そして通信部 2 は、表記データと読みデータを処理部 5 へ渡す。

【0017】

なお、語句の表記は、例えば、漢字と仮名の組み合わせ、または漢字のみで表記され、一方、語句の読みは、例えば、仮名で表記される。なお、以下では、説明の便宜上、語句の表記を漢字仮名混じり文で表記し、一方、語句の読みを片仮名で表記する。また、拗音及び促音を、それぞれ、小文字で表記する。

30

また、取得される読みデータでは、拗音及び促音は、直音と区別されずに表記されるものとする。例えば、取得される読みデータでは、拗音及び促音は小文字の仮名で表記されず、直音と同様に大文字の仮名で表記される。

また、表記データ及び読みデータは、どのようなファイル形式で表されていてもよく、例えば、表記データ及び読みデータはテキストファイルで表される。

【0018】

また通信部 2 は、処理部 5 から受け取った、語句の表記データ及び拗音または促音についての修正がなされた読みデータを、その語句が登録される辞書を記憶した装置へ通信ネットワークを介して出力してもよい。

40

【0019】

記憶部 3 は、例えば、半導体メモリ回路、磁気記憶装置または光記憶装置のうちの少なくとも一つを有する。そして記憶部 3 は、処理部 5 で用いられる各種コンピュータプログラム、及び、拗促音判定処理を含む辞書登録処理に用いられる各種のデータを記憶する。例えば、記憶部 3 は、様々な漢字の表記及び読みと、その読みが音読みか訓読みかを表す音訓種別情報が登録された単漢字辞書を記憶する。

【0020】

また記憶部 3 は、辞書に登録する対象となる語句の表記データと読みデータを記憶して

50

いてもよい。さらに、記憶部 3 は、二つ以上の漢字が一つの単語として並べて使用される共起確率を表す共起確率データベースを記憶していてもよい。あるいは、記憶部 3 は、そのような共起確率データベースを決定するために利用される検索コーパスを記憶していてもよい。なお、共起確率データベース及び検索コーパスの詳細については後述する。さらに、記憶部 3 は、語句の表記及び読みを登録すべき辞書を記憶していてもよい。語句の表記及び読みが登録される辞書は、例えば、音声合成または音声認識において使用される単語辞書であってもよい。

#### 【 0 0 2 1 】

表示部 4 は、例えば、液晶ディスプレイといった表示装置を有する。そして表示部 4 は、例えば、登録対象の語句の表記データと、拗音及び促音の表記の修正がなされたその語句の読みデータとを表示する。

10

#### 【 0 0 2 2 】

処理部 5 は、一つまたは複数のプロセッサと、メモリ回路と、周辺回路とを有する。そして処理部 5 は、登録対象の語句の読みデータに含まれる、拗音または促音となり得る音が拗音または促音であるか否かを判定する。そして処理部 5 は、その音が拗音または促音である場合、読みデータに含まれる、語句の読みの拗音または促音と判定された音の表記を修正した上で、その語句の読みを、対応する表記とともに辞書に登録する。

#### 【 0 0 2 3 】

図 2 は、処理部 5 の機能ブロック図である。処理部 5 は、読み候補検出部 1 1 と、拗音候補抽出部 1 2 と、対応付け部 1 3 と、判定部 1 4 と、登録部 1 5 とを有する。

20

処理部 5 が有するこれらの各部は、例えば、処理部 5 が有するプロセッサ上で動作するコンピュータプログラムにより実現される機能モジュールである。あるいは、処理部 5 が有するこれらの各部は、その各部の機能を実現する一つの集積回路として辞書登録装置 1 に実装されてもよい。

#### 【 0 0 2 4 】

読み候補検出部 1 1 は、登録対象の語句を、その語句の表記に含まれる漢字ごとに分解する。そして読み候補検出部 1 1 は、単漢字辞書を参照して、その語句の表記に含まれる各漢字の読みを、その漢字の読み候補として検出する。

#### 【 0 0 2 5 】

例えば、登録対象の語句の表記が「滅菌」である場合、漢字「滅」と「菌」が含まれている。単漢字辞書に登録されている漢字「滅」の読みが「メツ」及び「メチ」であり、漢字「菌」の読みが「キン」及び「ゴン」とする。この場合、読み候補検出部 1 1 は、漢字「滅」の読み候補として、「メツ」及び「メチ」を検出し、漢字「菌」の読み候補として、「キン」及び「ゴン」を検出する。さらに、読み候補検出部 1 1 は、単漢字辞書を参照して、各読み候補の音訓種別情報を取得する。

30

#### 【 0 0 2 6 】

読み候補検出部 1 1 は、登録対象の語句に含まれる各漢字の読み候補及びその読み候補の音訓種別情報を対応付け部 1 3 へ出力する。

#### 【 0 0 2 7 】

拗音候補抽出部 1 2 は、登録対象の語句の読みデータから、拗音または促音となり得る音を抽出し、抽出した音が、拗音または促音が含まれる可能性がある特定の音の並びに含まれる場合に、その抽出した音を拗音または促音の候補とする。本実施形態では、拗音となり得る音の候補を拗音候補とし、促音となり得る音の候補を促音候補とする。

40

#### 【 0 0 2 8 】

例えば、拗音候補抽出部 1 2 は、登録対象の語句の読みから、拗音となり得る音「ヤ」、「ユ」、「ヨ」を抽出する。そして拗音候補抽出部 1 2 は、抽出した音の直前の音が複数の特定の音の何れかである場合、抽出した音を拗音候補とする。

#### 【 0 0 2 9 】

図 3 は、拗音を含む可能性がある音の並びのパターンを示す図である。リスト 3 0 0 は、拗音となり得る音の一覧を示す。具体的には、リスト 3 0 0 には、「ヤ」、「ユ」、「

50

ヨ」が含まれる。また、リスト 3 1 0 は、直後の音がリスト 3 0 0 に含まれる音である場合に、リスト 3 0 0 に含まれる音が拗音となる可能性がある音の一覧を示す。具体的には、リスト 3 1 0 には、「キノギ」、「シノジ」、「チノヂ」、「ニ」、「ヒノビノピ」、「ミ」、「リ」が含まれる。

#### 【 0 0 3 0 】

例えば、読みが「ジユウ」であれば、拗音となり得る音「ユ」の直前の音が、リスト 3 1 0 に含まれる音「ジ」であるので、拗促音候補抽出部 1 2 は、読み「ジユウ」に含まれる音「ユ」を拗音候補とする。また、読みが「キヨウイク」であれば、拗音となり得る音「ヨ」の直前の音が、リスト 3 1 0 に含まれる音「キ」であるので、拗促音候補抽出部 1 2 は、読み「キヨウイク」に含まれる音「ヨ」を拗音候補とする。

10

一方、読み「イヤク」には、拗音となり得る音「ヤ」が含まれている。しかし、「ヤ」の直前の音「イ」は、リスト 3 1 0 に含まれていない。そのため、拗促音候補抽出部 1 2 は、読み「イヤク」に含まれる音「ヤ」を拗音候補にしない。

#### 【 0 0 3 1 】

また、拗促音候補抽出部 1 2 は、登録対象の語句の読みから、促音となり得る音「ツ」を抽出する。そして拗促音候補抽出部 1 2 は、抽出した音の直前の音が特定の音の何れかであり、かつ、抽出した音の直後の音が複数の特定の音の何れかである場合、抽出した音を促音候補とする。

#### 【 0 0 3 2 】

図 4 は、促音を含む可能性がある音の並びのパターンを示す図である。リスト 4 0 0 は、促音となり得る音の一覧を示す。具体的には、リスト 4 0 0 には、「ツ」が含まれる。また、リスト 4 1 0 は、促音の直前に位置する可能性がある音の一覧を表す。一方、リスト 4 2 0 は、促音の直後に位置する可能性がある音の一覧を表す。すなわち、リスト 4 0 0 に含まれる音「ツ」の直前に、リスト 4 1 0 に含まれる何れかの音が位置し、かつ、音「ツ」の直後に、リスト 4 2 0 に含まれる何れかの音が位置する場合に、音「ツ」は促音候補となる。

20

#### 【 0 0 3 3 】

具体的には、リスト 4 1 0 には、「ヲ」及び「ン」以外の全ての音が含まれ、リスト 4 2 0 には、「カ行」、「サ行」、「タ行」及び「パ行」の音が含まれる。

#### 【 0 0 3 4 】

例えば、読みが「ガツコウ」であれば、促音となり得る音「ツ」の直前の音が、リスト 4 1 0 に含まれる音「ガ」であり、かつ、「ツ」の直後の音「コ」が、リスト 4 2 0 に含まれる。そこで、拗促音候補抽出部 1 2 は、読み「ガツコウ」に含まれる音「ツ」を促音候補とする。

30

一方、読み「フツウ」には、促音となり得る音「ツ」が含まれている。しかし、「ツ」の直後の音「ウ」は、リスト 4 2 0 に含まれていない。そのため、拗促音候補抽出部 1 2 は、読み「フツウ」に含まれる音「ツ」を促音候補にしない。

#### 【 0 0 3 5 】

拗促音候補抽出部 1 2 は、拗音候補及び促音候補を対応付け部 1 3 及び判定部 1 4 へ出力する。

40

#### 【 0 0 3 6 】

対応付け部 1 3 は、登録対象の語句に含まれる各漢字ごとに、その漢字の読み候補のうち、その語句についての読みと最も一致する候補を特定し、特定された候補に対応する語句の読みの部分を、その漢字の読みとする。

#### 【 0 0 3 7 】

本実施形態では、対応付け部 1 3 は、対象語句の読みに含まれる促音候補である音「ツ」を、音「ウ/ク/ス/チ」の何れかと置換した上で、各漢字の読みと比較する。なお、促音候補と置換する音「ウ/ク/ス/チ」は、ある漢字の読みで用いられ、かつ、促音に置き換えられる可能性がある音である。

さらに、対応付け部 1 3 は、漢字の読み候補に含まれる拗音または促音については、直

50

音と一致するとみなして、語句の読みと漢字の読み候補を比較する。例えば、漢字の読みの候補において拗音及び促音が小文字で表記され、直音が大きくて表記されている場合、対応付け部 13 は、語句の読みと漢字の読み候補とを小文字と大文字を区別せずに比較する。このような置換などを行うことにより、対応付け部 13 は、語句の中で用いられることにより読みが変化した漢字についても、語句の読みとその漢字の読みとを対応付けることができる。そのため、対応付け部 13 は、語句の読みのうち、その語句に含まれる各漢字に対応する部分を正確に決定できる。

#### 【0038】

対応付け部 13 は、対象語句の先頭から順に、その語句に含まれる漢字の読みの候補とその語句の読みとを比較し、漢字の読みの候補のうちで最も一致する候補を、対象語句の読みのうちのその漢字に対応する部分とする。

10

あるいは、対応付け部 13 は、対象語句に含まれる各漢字の読みの候補の組み合わせのそれぞれと、その語句の読みとを、動的計画法(Dynamic Programming, DP)マッチングにより対応付けることで、最も一致する組み合わせを特定してもよい。そして対応付け部 13 は、その組み合わせにおいて各漢字の読みの候補と対応付けられた、対象語句の読みの部分を、その漢字に対応する読みとしてもよい。

#### 【0039】

対応付け部 13 は、登録対象の語句の読みと、その語句に含まれる各漢字との対応関係を表す情報、例えば、漢字ごとの読みの境界の位置を判定部 14 へ通知する。

#### 【0040】

20

判定部 14 は、登録対象の語句の読みに含まれる拗音候補と対応するその語句に含まれる漢字の読みにおける拗音候補の位置、及び、その漢字の読みの音訓種別に応じて、拗音候補が拗音か否かを判定する。同様に、判定部 14 は、登録対象の語句の読みに含まれる促音候補と対応するその語句に含まれる漢字の読みにおける促音候補の位置、及び、その漢字の読みの音訓種別に応じて、促音候補が促音か否かを判定する。

#### 【0041】

本実施形態では、判定部 14 は、拗音候補が、対応する漢字の読みにおける先頭に位置するか、対応する漢字の読みが訓読みの場合、拗音候補は拗音でないと判定する。一方、判定部 14 は、拗音候補が、対応する漢字の読みにおける先頭以外に位置し、かつ、対応する漢字の読みが音読みであれば、拗音候補は拗音であると判定する。

30

#### 【0042】

図 5 は、拗音候補と対応漢字の読みとの位置関係の一例を示す図である。図 5 には、登録対象語句に含まれる漢字とその漢字に対応する読みとの 3 種類の組み合わせ 501 ~ 503 が示されている。組み合わせ 501 には、漢字「銃」に対して読み「ジユウ」が対応付けられており、このうち、音「ユ」が拗音候補である。そしてこの拗音候補は、対応する漢字「銃」の読みの先頭ではなく、かつ、読み「ジユウ」は、音読みである。したがって、判定部 14 は、拗音候補「ユ」を拗音とする。

#### 【0043】

また、組み合わせ 502 には、漢字「教」に対して読み「キヨウ」が対応付けられており、かつ、漢字「育」に対して読み「イク」が対応付けられている。このうち、音「ヨ」が拗音候補である。そしてこの拗音候補は、対応する漢字「教」の読みの先頭ではなく、かつ、読み「キヨウ」は、音読みである。したがって、判定部 14 は、拗音候補「ヨ」を拗音とする。

40

#### 【0044】

一方、組み合わせ 503 には、漢字「自」に対して読み「ジ」が対応付けられており、かつ、漢字「由」に対して読み「ユウ」が対応付けられている。このうち、音「ユ」が拗音候補である。そしてこの拗音候補は、対応する漢字「由」の読みの先頭に位置している。したがって、判定部 14 は、拗音候補「ユ」は、拗音でないと判定する。

#### 【0045】

また、判定部 14 は、促音候補が、対応する漢字の読みにおける末尾以外に位置するか

50

、単語境界に接する位置にあるか、あるいは、対応する漢字の読みが訓読みである場合、促音候補は促音でないと判定する。一方、判定部14は、促音候補が、対応する漢字の読みにおける末尾に位置し、単語境界に接しておらず、かつ、対応する漢字の読みが音読みであれば、促音候補は促音であると判定する。なお、単語境界は、互いに独立して用いられる単語間の境界である。

#### 【0046】

図6は、促音候補と対応漢字の読み及び単語境界との位置関係の一例を示す図である。図6には、登録対象語句に含まれる漢字とその漢字に対応する読みとの3種類の組み合わせ601～603が示されている。組み合わせ601には、二つの漢字「学校」に対して読み「ガツコウ」が対応付けられており、このうち、音「ツ」が促音候補である。そしてこの促音候補は、対応する漢字「学」の読みの末尾に位置し、かつ、漢字「学」「校」の間は単語境界ではない。さらに、読み「ガツコウ」は、音読みである。したがって、判定部14は、促音候補「ツ」を促音とする。

10

#### 【0047】

また、組み合わせ602には、二つの漢字「大月」に対して読み「オオツキ」が対応付けられており、このうち、音「ツ」が促音候補である。そしてこの促音候補は、対応する漢字「月」の読みの先頭に位置するので、判定部14は、促音候補「ツ」を促音ではないと判定する。

#### 【0048】

一方、組み合わせ603には、二つの漢字「脱税」に対して読み「ダツゼイ」が対応付けられている。このうち、音「ツ」が促音候補である。この拗音候補は、対応する漢字「脱」の読みの末尾に位置しているものの、漢字「脱」「税」の間は単語境界である。したがって、判定部14は、促音候補「ツ」は、促音でないと判定する。

20

#### 【0049】

判定部14は、二つの漢字の境界が単語境界に相当するか否かを判定するために、例えば、共起確率データベースを参照する。

#### 【0050】

図7は、共起確率データベースの一例を示す図である。共起確率データベース700では、上端の各欄に、二つの漢字の並びのうちの前側に位置する漢字が示され、左端の各欄に、二つの漢字の並びのうちの後側に位置する漢字が示される。そして共起確率データベース700のその他の各欄には、その欄と同じ列の上端に示された漢字に後続して、その欄と同じ行の左端に示された漢字が表れる共起確率を表す。例えば、この例では、漢字「滅」に後続して漢字「菌」が表れる共起確率は0.75である。

30

なお、共起確率データベースは、二つの漢字の組み合わせの共起確率を示すものに限られず、一方または両方が2以上の漢字を含む語句の組み合わせの共起確率を示していてもよい。

#### 【0051】

判定部14は、促音候補が読みの末尾に位置する漢字と、その次の漢字の組み合わせに対する共起確率を、共起確率データベースを参照して決定する。そして判定部14は、その共起確率が所定の閾値未満である場合、その二つの漢字の境界は単語境界であると判定し、一方、共起確率が所定の閾値以上である場合、その二つの漢字の境界は単語境界でないと判定する。なお、所定の閾値は、例えば、0.6に設定される。

40

#### 【0052】

さらに、判定部14は、登録対象の語句の先頭及び末尾を単語境界としてもよい。

#### 【0053】

また、判定部14は、単語境界の位置を他の方法に従って決定してもよい。例えば、判定部14は、登録対象の語句に対して形態素解析を適用して、その語句を形態素ごとに分割してもよい。そして判定部14は、形態素間の境界を単語境界としてもよい。

#### 【0054】

あるいは、判定部14は、接尾語と、その直前の漢字との間を単語境界としてもよい。

50

同様に、判定部 14 は、接頭語と、その直後の漢字との間を単語境界としてもよい。例えば、「初」あるいは「活」といった接頭語が付された語句（ただし、は任意の漢字を表す）では、判定部 14 は、「初」及び「活」が接頭語なので、「初」または「活」とその直後の漢字の間を単語境界としてもよい。また、「期」といった接尾語が付された語句（ただし、は任意の漢字を表す）では、判定部 14 は、「期」が接尾語なので、「期」とその直前の漢字の間を単語境界としてもよい。

【0055】

あるいはまた、判定部 14 は、人名の一部として用いられる漢字と、その直前又は直後の漢字との境界を、単語境界としてもよい。人名の一部として用いられる漢字は、例えば、「子」、「夫」、「男」、「雄」、あるいは、「美」である。例えば、「子」とい

10

【0056】

判定部 14 は、読みデータにおいて、拗音と判定された拗音候補を、拗音に補正する。同様に、判定部 14 は、読みデータにおいて、促音と判定された促音候補を、促音に補正する。例えば、読みデータがテキストデータである場合、判定部 14 は、拗音及び促音を、小文字表記に補正する。判定部 14 は、登録対象の語句の表記データと、拗音候補及び促音候補のうち、拗音及び促音と判定された音が補正されたその語句の読みデータ（以下、修正済み読みデータと呼ぶ）とを、表示部 4 に表示させる。

【0057】

20

登録部 15 は、登録対象の語句の表記データと、修正済み読みデータとを、記憶部 3 に記憶されている辞書に登録する。なお、語句を登録すべき辞書が、辞書登録装置 1 と通信ネットワークを介して接続される他の装置に記憶されている場合、登録部 15 は、登録対象の語句の表記データと、修正済み読みデータとを、通信部 2 を介して他の装置へ送信してもよい。

【0058】

図 8 は、辞書登録装置 1 の処理部 5 によって実行される、拗促音判定処理を含む辞書登録処理の動作フローチャートである。なお、以下に示す動作フローチャートにおいて、ステップ S101 ~ S108 の処理が拗促音判定処理に相当する。処理部 5 は、通信部 2 を介して登録対象語句の表記データと読みデータを取得する度に、以下の動作フローチャートに従って、辞書登録処理を実行する。

30

【0059】

読み候補検出部 11 は、単漢字辞書を参照して、登録対象の語句に含まれる漢字ごとに、その漢字の読みを読み候補として検出する（ステップ S101）。読み候補検出部 11 は、登録対象の語句に含まれる各漢字の読み候補及びその読み候補の音訓種別情報を対応付け部 13 へ出力する。

【0060】

拗促音候補抽出部 12 は、登録対象の語句の読みに含まれる拗音となり得る音のうち、拗音を含む可能性がある特定の音の並びに含まれる音を拗音候補として抽出する（ステップ S102）。また、拗促音候補抽出部 12 は、登録対象の語句の読みに含まれる促音となり得る音のうち、促音を含む可能性がある特定の音の並びに含まれる音を促音候補として抽出する（ステップ S103）。拗促音候補抽出部 12 は、抽出された拗音候補及び促音候補を対応付け部 13 及び判定部 14 へ出力する。

40

【0061】

対応付け部 13 は、登録対象の語句に含まれる漢字ごとに、その漢字の読みの候補と、その語句についての読みとを比較して、その語句の読みのうち、各漢字に対応する部分を特定する（ステップ S104）。対応付け部 13 は、登録対象の語句の読みと各漢字との対応関係を表す情報を判定部 14 へ通知する。

【0062】

判定部 14 は、拗音候補が、対応する漢字の読みにおける先頭以外に位置し、かつ、対

50

応する漢字の読みが音読みか否か判定する（ステップS105）。拗音候補が、対応する漢字の読みにおける先頭以外に位置し、かつ、対応する漢字の読みが音読みであれば（ステップS105 - Yes）、判定部14は、拗音候補は拗音であると判定する。そして判定部14は、読みデータにおいて、その拗音候補を拗音に補正する（ステップS106）。

一方、拗音候補が、対応する漢字の読みにおける先頭に位置するか、あるいは、対応する漢字の読みが訓読みであれば（ステップS105 - No）、判定部14は、拗音候補は拗音でないと判定する。

#### 【0063】

ステップS106の後、あるいは、判定部14が拗音候補は拗音でないと判定した後、判定部14は、促音候補が、対応する漢字の読みにおける末尾に位置し、単語境界と接しておらず、かつ、対応する漢字の読みが音読みか否か判定する（ステップS107）。促音候補が、対応する漢字の読みにおける末尾に位置し、単語境界と接しておらず、かつ、対応する漢字の読みが音読みである場合（ステップS107 - Yes）、判定部14は、促音候補は促音であると判定する。そして判定部14は、読みデータにおいて、その促音候補を促音に補正する（ステップS108）。

一方、促音候補が、対応する漢字の読みの末尾以外に位置するか、単語境界に接しているか、あるいは、対応する漢字の読みが訓読みであれば（ステップS107 - No）、判定部14は、促音候補は促音でないと判定する。

#### 【0064】

ステップS108の後、あるいは、判定部14が促音候補は促音でないと判定した後、登録部15は、登録対象の語句の表記データと修正済み読みデータとを辞書に登録する（ステップS109）。その後、処理部5は、辞書登録処理を終了する。

なお、処理部5は、ステップS101の処理と、ステップS102及びS103の処理の順序を入れ替えてもよい。また処理部5は、ステップS105及びS106の処理と、ステップS107及びS108の処理の順序を入れ替えてもよい。

#### 【0065】

以上に説明してきたように、この辞書登録装置は、拗音または促音を含む可能性がある特定の音の並びに含まれる拗音候補または促音候補が、拗音または促音となり得る、対応する漢字の読みの位置にあるか否か、及び対応する漢字の読みの音訓種別を調べる。これにより、この辞書登録装置は、拗音候補及び促音候補が、それぞれ、拗音であるか、または促音であるかを適切に判定できる。そのため、この辞書登録装置は、作業者が手作業で読みデータ中の拗音及び促音の表記を修正しなくても、自動的に拗音及び促音の表記を修正した上で、語句の表記データと読みデータを辞書に登録できるので、辞書作成の際の工数を削減できる。

#### 【0066】

なお、処理部5は、促音候補に対応する漢字とその直後の漢字の組み合わせについての共起確率を、検索用コーパスに基づいて求めてよい。

#### 【0067】

図9は、変形例による辞書登録装置1の処理部5の機能ブロック図である。この変形例では、処理部5は、読み候補検出部11と、拗促音候補抽出部12と、対応付け部13と、判定部14と、登録部15と、共起確率算出部16とを有する。この変形例による処理部5は、図2に示された処理部5と比較して、共起確率算出部16を有する点で異なる。そこで、以下では、共起確率算出部16及びその関連部分について説明する。

#### 【0068】

判定部14は、促音候補が対応する漢字の読みの末尾に位置する場合、その漢字及び直後の漢字の組み合わせを、共起確率の算出対象の漢字の組み合わせとして、共起確率算出部16へ通知する。

#### 【0069】

共起確率算出部16は、共起確率の算出対象の漢字の組み合わせを含む検索用コーパス

10

20

30

40

50

を取得する。共起確率算出部 16 は、検索用コーパスとして、例えば、辞書登録装置 1 と通信ネットワークを介して接続されている様々なWebサイト上で公開されているWebページに含まれる、漢字仮名混じり文のテキストデータを、利用できる。あるいは、検索用コーパスとして、例えば、製品のマニュアル、議事録、メールなどに含まれる漢字仮名混じり文のテキストデータを利用できる。共起確率算出部 16 は、例えば、これらの検索用コーパスを、記憶部 3 から読み込む。あるいは、共起確率算出部 16 は、通信部 2 を介して、辞書登録装置 1 と通信ネットワークを介して接続されている他の機器から、検索用コーパスを取得する。

#### 【0070】

共起確率算出部 16 は、例えば、検索用コーパスに含まれるテキストデータに含まれる、共起確率の算出対象の漢字の組み合わせの個数と、その組み合わせに含まれる各漢字の個数を、それぞれカウントする。そして共起確率算出部 16 は、次式に従って、共起確率 CoP を算出する。

$$\text{CoP} = 2N_{12} / (N_1 + N_2)$$

ここで、 $N_1$  は、共起確率の算出対象の漢字の組み合わせに含まれる先頭の漢字の個数を表し、 $N_2$  は、共起確率の算出対象の漢字の組み合わせに含まれる後続の漢字の個数を表す。そして  $N_{12}$  は、その漢字の組み合わせの個数を表す。

例えば、共起確率算出対象となる漢字の組み合わせが「滅菌」であり、検索用コーパスに、その組み合わせ「滅菌」が 3 個ふくまれており、個々の漢字「滅」及び「菌」がそれぞれ 4 個含まれているとする。この場合、共起確率 CoP は、 $2 \times 3 / (4 + 4) = 0.75$  となる。

また、共起確率算出部 16 は、別の共起確率の算出方法として、Web 検索サービスの検索件数を利用して共起確率を算出してもよい。Web 検索サービスを用いる場合、共起確率算出部 16 は、次式で共起確率 CoP を算出する。

$$\text{CoP} = (\text{単語の組み合わせの検索件数}) / (\text{単語 1 と単語 2 の and 条件での検索件数})$$

例えば、Web 検索サービスで「滅菌」と検索した場合と、「滅 and 菌」で検索した場合の検索結果が、それぞれ 1,270,000 件、1,750,000 件であるとする、共起確率 CoP は、 $1,270,000 / 1,750,000 = 0.72$  となる。

#### 【0071】

共起確率算出部 16 は、算出された共起確率を判定部 14 へ通知する。そして判定部 14 は、通知された共起確率に基づいて、促音候補が単語境界に接しているか否かを判定すればよい。

また、共起確率算出部 16 は、共起確率データベースに、その算出対象となった漢字の組み合わせ及び共起確率を追加してもよい。

#### 【0072】

この変形例によれば、辞書登録装置 1 は、登録対象の語句から得られた促音候補に関連する漢字の組み合わせについての共起確率が事前に分かっているなくても、その組み合わせについての共起確率を算出できる。そのため、辞書登録装置 1 は、予め多数の漢字の組み合わせに対する共起確率を表した大規模な共起確率データベースを持たなくても、促音候補が単語境界に接しているか否かを判定できるので、その促音候補が促音か否かを適切に判定できる。

#### 【0073】

また、他の変形例によれば、記憶部 3 は、促音候補に対応する漢字及びその前後の漢字の組み合わせに関して、促音候補を促音とする例外リストを記憶していてもよい。そして判定部 14 は、促音候補に対応する漢字及びその前後の漢字の組み合わせが例外リストに登録されている場合、促音候補が単語境界に接しているか否か、及び、促音候補に対応する漢字の読みの音訓種別に関わらず、促音候補と促音としてもよい。

例えば、「勝平（カッペイ）」といった人名、または「鳥取（トットリ）」といった地名のように、固有名詞では、漢字の読みが訓読みであっても、その読みの中に促音が含まれることがある。そこでこのような例外的な漢字の組み合わせを例外リストとして予め登録しておくことで、判定部 14 は、より正確に促音候補を促音か否かを判定できる。

## 【 0 0 7 4 】

さらに他の変形例によれば、辞書登録装置は、複数の語句を含む文字列の表記データとその文字列の読みデータとを含むデータベースを、例えば、通信部 2 を介して取得してもよい。この場合、処理部 5 は、表記データに表された文字列に対して、例えば、形態素解析を適用することにより、その文字列を語句ごとに分解する。そして処理部 5 は、得られた語句のうち、辞書に登録されていない語句に対して、上記の実施形態または変形例による辞書登録処理を実行してもよい。これにより、辞書登録装置は、一つのデータベースから、複数の語句を自動的に辞書に登録することができる。

## 【 0 0 7 5 】

さらに、上記の実施形態または変形例による辞書登録装置は、登録対象の語句の表記データ及び読みデータを、それらのデータが記録された磁気記録媒体あるいは光記録媒体といった記録媒体から読み込んでもよい。この場合には、辞書登録装置は、取得部として、そのような記録媒体のアクセス装置を有していてもよい。

## 【 0 0 7 6 】

さらに、上記の実施形態または変形例による辞書登録装置の処理部が有する各機能をコンピュータに実現させるコンピュータプログラムは、コンピュータによって読み取り可能な媒体、例えば、磁気記録媒体、光記録媒体または半導体メモリに記録された形で提供されてもよい。

## 【 0 0 7 7 】

ここに挙げられた全ての例及び特定の用語は、読者が、本発明及び当該技術の促進に対する本発明者により寄与された概念を理解することを助ける、教示的な目的において意図されたものであり、本発明の優位性及び劣等性を示すことに関する、本明細書の如何なる例の構成、そのような特定の挙げられた例及び条件に限定しないように解釈されるべきものである。本発明の実施形態は詳細に説明されているが、本発明の精神及び範囲から外れることなく、様々な変更、置換及び修正をこれに加えることが可能であることを理解されたい。

## 【 0 0 7 8 】

以上説明した実施形態及びその変形例に関し、更に以下の付記を開示する。

## ( 付 記 1 )

少なくとも一つの漢字を含む語句の表記を表す表記データ及び当該語句の読みを拗音及び促音と直音とを区別せずに表す読みデータを取得する取得部と、

複数の漢字のそれぞれの読み及び当該読みが音読みか訓読みかを表す音訓種別が登録された単漢字辞書を記憶する記憶部と、

前記単漢字辞書を参照して、前記表記データに表された前記語句に含まれる前記少なくとも一つの漢字のそれぞれの読みの候補を検出する読み候補検出部と、

前記読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出する拗促音候補抽出部と、

前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定し、該特定された候補に対応する前記語句の読みの部分を、当該漢字の読みとする対応付け部と、

前記語句に含まれる前記少なくとも一つの漢字のうちの前記拗促音候補と対応する第 1 の漢字の読みにおける前記拗促音候補の位置、及び、前記第 1 の漢字の前記読みの音訓種別に応じて前記拗促音候補を拗音及び促音の何れかか否か判定する判定部と、  
を有する拗促音判定装置。

## ( 付 記 2 )

前記判定部は、前記拗促音候補が、促音となり得る音であり、前記第 1 の漢字の前記読みの末尾に位置し、かつ、互いに独立して用いられる単語間の境界と接しておらず、かつ、前記第 1 の漢字の前記読みが音読みである場合、前記拗促音候補を促音とする、付記 1 に記載の拗促音判定装置。

10

20

30

40

50

## (付記 3)

前記判定部は、前記第 1 の漢字と当該第 1 の漢字に後続する第 2 の漢字とが組み合わせて使用される共起確率が所定の閾値未満である場合、前記第 1 の漢字と前記第 2 の漢字の境界は前記単語間の境界であると判定する、付記 2 に記載の拗促音判定装置。

## (付記 4)

前記判定部は、前記拗促音候補が、拗音となり得る音であり、かつ、前記第 1 の漢字の読みの先頭以外に位置し、かつ、前記第 1 の漢字の前記読みが音読みである場合、前記拗促音候補を拗音とする、付記 1 ~ 3 の何れか一項に記載の拗促音判定装置。

## (付記 5)

前記対応付け部は、前記拗促音候補のうち、促音となり得る音を、促音に置換される可能性がある音に置換して、前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定する、付記 1 ~ 4 の何れか一項に記載の拗促音判定装置。

10

## (付記 6)

前記対応付け部は、前記漢字の読みの候補に含まれる拗音及び促音が、当該拗音及び促音を直音化した音と一致するとみなして、前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定する、付記 1 ~ 5 の何れか一項に記載の拗促音判定装置。

## (付記 7)

前記表記データと、促音または拗音と判定された前記拗促音候補を促音または拗音を表す表記に修正した前記読みデータとを、語句の表記と読みの関係を表す辞書に登録する登録部をさらに有する、付記 1 ~ 6 の何れか一項に記載の拗促音判定装置。

20

## (付記 8)

複数の漢字のそれぞれの読み及び当該読みが音読みか訓読みかを表す音訓種別が登録された単漢字辞書を参照して、少なくとも一つの漢字を含む語句の表記を表す表記データに表された前記語句に含まれる前記少なくとも一つの漢字のそれぞれの読みの候補を検出し、

拗音及び促音と直音とを区別せずに表す読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出し、

30

前記読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出し、

前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定し、該特定された候補に対応する前記語句の読みの部分を、当該漢字の読みとし、

前記語句に含まれる前記少なくとも一つの漢字のうちの前記拗促音候補と対応する第 1 の漢字の読みにおける前記拗促音候補の位置、及び、前記第 1 の漢字の前記読みの音訓種別に応じて前記拗促音候補を拗音及び促音の何れか否か判定する、

ことを含む拗促音判定方法。

40

## (付記 9)

複数の漢字のそれぞれの読み及び当該読みが音読みか訓読みかを表す音訓種別が登録された単漢字辞書を参照して、少なくとも一つの漢字を含む語句の表記を表す表記データに表された前記語句に含まれる前記少なくとも一つの漢字のそれぞれの読みの候補を検出し、

拗音及び促音と直音とを区別せずに表す読みデータに表された前記語句の読みに含まれる、促音または拗音となり得る音のうち、促音または拗音が含まれる可能性がある音の並びに含まれる音を拗促音候補として抽出し、

前記語句に含まれる前記少なくとも一つの漢字のそれぞれについて、当該漢字の読みの候補のうち、前記語句の読みと最も一致する候補を特定し、該特定された候補に対応する

50

前記語句の読みの部分を、当該漢字の読みとし、

前記語句に含まれる前記少なくとも一つの漢字のうちの前記拗促音候補と対応する第1の漢字の読みにおける前記拗促音候補の位置、及び、前記第1の漢字の前記読みの音訓種別に応じて前記拗促音候補を拗音及び促音の何れかか否か判定すること、

ことをコンピュータに実行させるための拗促音判定用コンピュータプログラム。

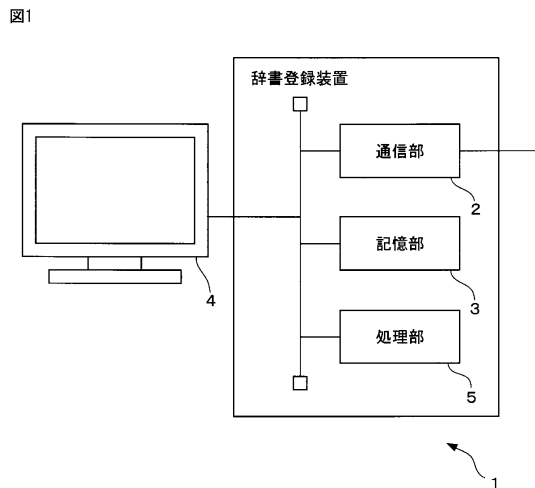
【符号の説明】

【0079】

- 1 辞書登録装置（拗促音判定装置）
- 2 通信部
- 3 記憶部
- 4 表示部
- 5 処理部
- 11 読み候補検出部
- 12 拗促音候補抽出部
- 13 対応付け部
- 14 判定部
- 15 登録部
- 16 共起確率算出部

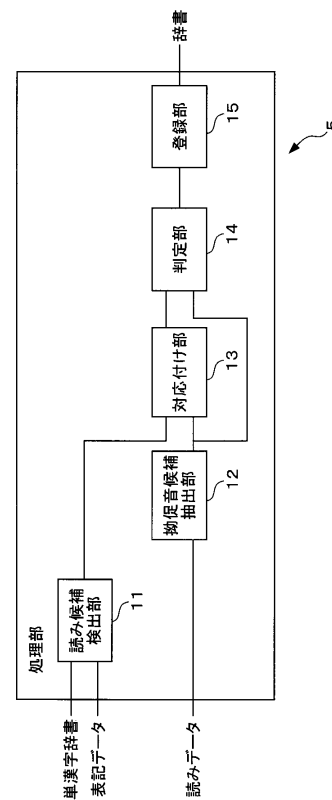
10

【図1】



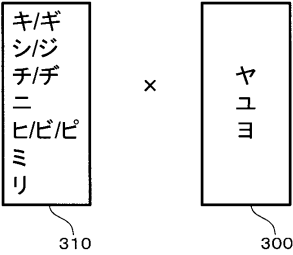
【図2】

図2



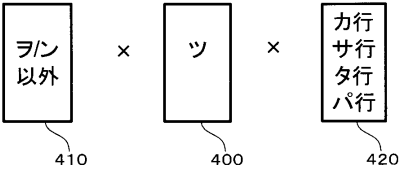
【図 3】

図3



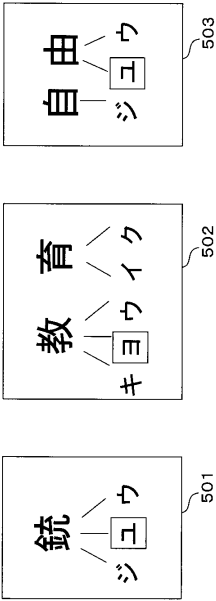
【図 4】

図4



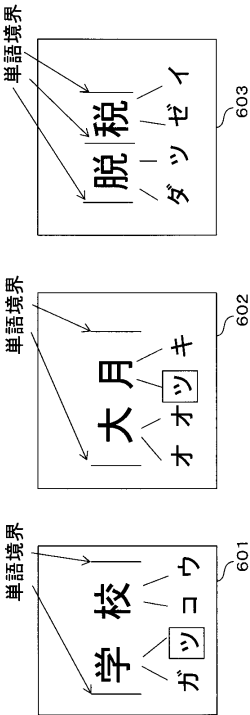
【図 5】

図5



【図 6】

図6



【図 7】

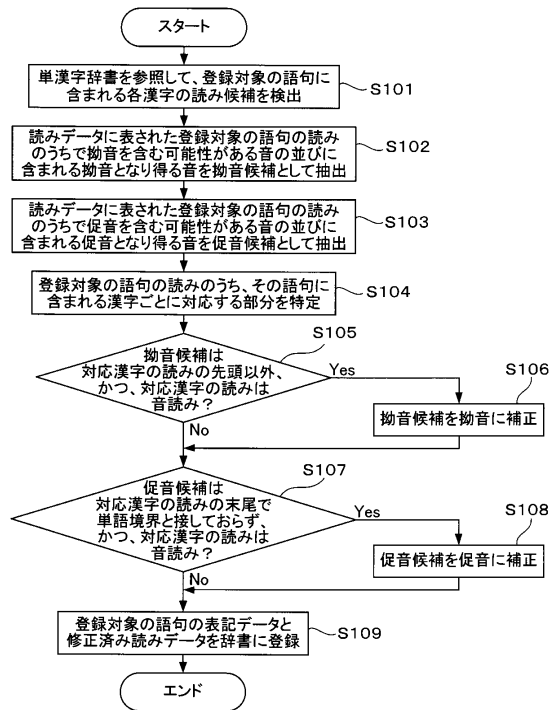
図7

漢字(前)			
	...	減	...
漢字(後)	...	...	...
固	...	0.75	...
...	...	...	...

700

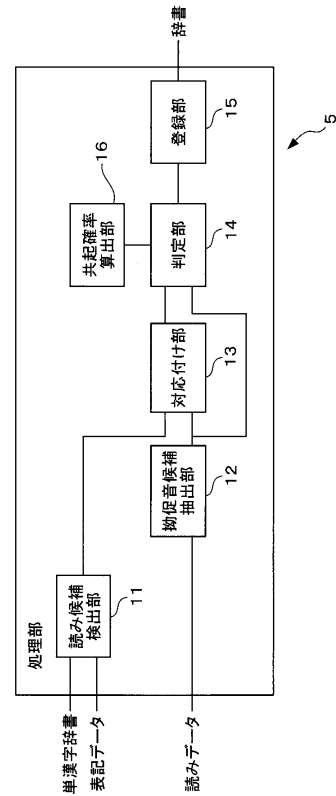
【図 8】

図8



【図 9】

図9



---

フロントページの続き

(56)参考文献 特開2004-139530(JP,A)  
特開2010-009294(JP,A)  
特開2013-041421(JP,A)  
特開平06-004520(JP,A)

(58)調査した分野(Int.Cl., DB名)  
G06F 17/20 - 17/28